GENOMICS AND EPIGENOMICS OF COMMON HUMAN METABOLIC AND HEART DISEASE

by

Michael Lewis Multhaup

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

February, 2015

Abstract

The field of epigenetics is rapidly becoming recognized as playing an essential part in explaining common human disease. Here we probe DNA methylation in diabetes mellitus and associated metabolic phenotypes and coronary heart disease. In a cohort from the Framingham Heart Study, we use epidemiological techniques to identify over 20,000 CpGs differentially methylated in coronary heart disease patients. In the other chapters, we use a functional approach to investigate the epigenetics of Type 2 Diabetes (T2D) and combine three lines of evidence – diet-induced epigenetic dysregulation in mouse, epigenetic conservation in humans, and evidence of T2D clinical risk – to identify genes implicated in T2D pathogenesis through epigenetic mechanisms related to obesity. We then replicate these results in adipose samples from lean and obese patients pre- and post-Roux-en-Y gastric bypass surgery, identifying regions where both the location and direction of methylation change is conserved. These regions overlap with 27 genomic locations with genetic T2D risk, only one of which was deemed significant by GWAS alone. Functional analysis of genes associated with these regions revealed five genes with novel roles in insulin resistance, demonstrating the potential general utility of this approach for complementing conventional human genetic studies by integrating cross-species epigenomics and clinical genetic risk.

Advisor: Andrew P. Feinberg

Reader: William G. Wong

Table of Contents

Acknowledgements

There have been many, many people who have had an impact on my graduate student career. Fellow students, friends, faculty, and family have all contributed to shape my journey through the Hopkins BCMB program. I would specifically like to thank my PI and mentor, Andy Feinberg. He has contributed to my growth both as a scientist and a mature adult. From him, I have not just learned about epigenetics, but also scholarship, thoughtfulness, critical thinking and professionalism. I could not have done this PhD without all the other members of the Feinberg Lab as well, who have contributed scientific insight and aid too numerous to mention. The members of the Epigenetic Center core facility have made this project possible, as we would have no array data without their expertise. My thesis committee has also been invaluable during this process, and I would like to thank William Wong, Natasha Zachara and Andrew Jaffe for their patience and scientific advice.

This work was supported by NIH Director Pioneer Award DP1 ES022579 to Andrew Feinberg, by NIDDK R01 grant DK084171 to William Wong, by the Novo-Nordisk Foundation and Stockholm County Council to Erik Naslund and Juleen Zierath, and by the European Research Council to Juleen Zierath. We thank Arni Runarsson for technical assistance. Michael Multhaup performed all of the molecular biology experiments and the Infinium 450k regression modeling, as well as performing many of the other statistical analyses in conjunction with Andrew Jaffe. Marcus Seldin performed the dietary manipulations of mice and isolated tissues under the oversight of Guang William Wong, with the help of Mehboob Hussain and his lab. Shelley Lei performed the functional studies of glucose uptake under the oversight of Guang William Wong. Andrew Jaffe performed statistical analysis. Alexander Drong and Mark McCarthy performed MAGENTA analysis. Henriette Kirschner, Erik Naslund and Juleen Zierath provided samples from the human obesity and RYGB cohort. Andrew Feinberg was responsible for the

design and oversight of all of the experiments and wrote papers with Michael Multhaup and considerable insight and assistance from Andrew Jaffe, Mark McCarthy, and Juleen Zierath.

List of Figures

List of Tables

Chapter 1

The history and current status of epigenetics

Epigenetics is typically defined as maintenance or reproduction of information through cellular division that occurs without changes in the primary DNA sequence. Originally coined by Conrad Waddington in 1942, the term “epigenetics” was originally used to mean the forces that shaped genotype into phenotype according to his “canalization” theory where the almost infinite potential chaos of the human genome was channeled into a much smaller number of recognizable phenotypes (Waddington, 1942). Many years later, Robin Holliday, contemplating his studies of drosophila development, more narrowly defined epigenetics as the manner in which the properties of genes change “during the development of the organism from fertilized egg to adult” (Holliday, 1987) and later as “the study of changes in gene expression, which occur in organisms with differentiated cells, and the mitotic inheritance of given patterns of gene expression” (Holliday, 1994). In this same paper, Holliday proposes a subcategory of epigenetics as a transmission of information “other than the DNA sequence itself.” This subcategory definition quickly gained traction, presumably because of its useful contrast to genetic inheritance, and has become common parlance today.

Quite a few biological mechanisms could fit within the umbrella of this definition. One could postulate that the anterior-posterior axis positioning in embryo development is epigenetic, as the positioning information (not genetic) gives rise to changes in gene expression that last the lifetime of the cells. Similarly, certain types of stem cell polarity arise from the local cellular environment (not genetic) and give rise to differential gene regulation patterns in daughter cells that are then passed on. In practice, however, the term “epigenetics” is most commonly used to refer to two systems: DNA

methylation (the covalent modification of DNA bases) and histone modification (the post-translational modifications of histone molecules or their substitution) (Dupont et al., 2009).

The covalent modification of DNA bases was first described in 1948, when a study attempting to perfect a method of using paper chromatography to separate out the five nucleotides found extra spectrophotomer absorption peaks, including one near cytosine that they termed “epicytosine” (Hotchkiss, 1948). Later, DNA modification was first postulated to have the potential to affect gene expression in a 1969 opinion piece that attempted to create a plausible explanation for the persistence of human memory and proposed long-term (covalent) modification of DNA in neurons leading to the alteration of gene expression within neurons – the “DNA ticketing theory of memory” (Griffith and Mahler, 1969). While this theory of human memory turned out to be incorrect, its mechanism may not have been too far afield; recent studies have found an enrichment of hydroxymethylation (5-hydroxymethylcytosine; 5-hmc) within certain areas of the brain (Wang et al., 2012).

In vertebrates, DNA methylation most commonly occurs at the cytosine in a cytosine-guanine dinucleotide (CG), though it has been found to occur infrequently in a few particular cell types on the cytosines in cytosine-alanine and cytosine-thymine sequences as well (Ramsahoye et al., 2000). This methylation occurs at the 5 position of the cytosine ring, counting towards the carboxyl group from the nitrogen between the amine and carboxyl groups. S-adenosylmethionine (SAM-CH_3) donates the methyl group, which then replaces the 5’-hydrogen atom to form 5-methylcytosine in a reaction catalyzed by a group of enzymes known as DNA methyltransferases (DNMTs).

There are three main DNMTs in mammals – DNMT1, DNMT3a and DNMT3b. All of these genes are essential, and full knockouts of these genes inevitably lead to mortality (though DNMT3a -/- mice can live to up to four weeks of age) (Okano et al., 1999). A fourth enzyme, DNMT2, was named for its close homology to DNMT1 but does not actually possess any DNA methyltransferase activity. It was recently revealed to actually be an RNA methyltransferase, responsible for the methylation of aspartic

acid transfer tRNA (Goll et al., 2006). All DNMTs have the same basic mechanism (Bestor, 2000). First, sequence recognition domains in the various enzymes make contact with bases in the major groove of the DNA double helix. Second, the target cytosine is removed from its usual base-pairing and inserted deep within the enzyme itself. Third, the enzyme makes a temporary covalent bond at the cytosine 4 position to create a more reactive 5 position. Fourth, the newly reactive 5 position strips a methyl group from S-adenosyl-methionine (SAM), the methyl donor. Fifth, the enzyme reforms the double bond between the 4 and 5 positions, releasing itself from the cytosine.

While all three DNMTs use this mechanism, their functions are very different. DNMT1 is the "maintenance" enzyme, responsible for the perpetuation of methylated cytosines through DNA replication and cell division. In DNA replication, newly synthesized DNA is created from unmethylated cytosines and therefore has no methylation. However, the semi-conservative nature of DNA replication means that every new strand of DNA will pair with a previously existing, methylated strand. This creates what is known as "hemi-methylated" DNA, where the methylated CGs on the parental strand are paired with unmethylated CGs on the daughter strand. DNMT1 specifically recognizes these hemi-methylated sites, binds to them, and methylates the unmethylated CG on the daughter strand. This leads to the theoretically perfect copying of methylation position information through cell division.

DNMT3a and 3b are known as the "de novo" methyltransferases, as they have an equal preference for unmethylated and hemi-methylated DNA (Okano et al., 1998). They are thought to be responsible for the placement of new methylation marks during development and when one cell type turns into another, as in hematopoiesis (Challen et al., 2014). The mechanism by which these methyltransferases are targeted to specific areas of the genome, however, is not yet fully understood. It is known that DNMTs can be attracted and/or activated by specific histone modifications (Li et al., 2011), raising the possibility that the histone code landscape could influence the DNA methylome. DNA methylation is also known to influence histone marks at specific genomic loci (Cedar and Bergman,

2009), which could potentially lead to a chicken-and-egg problem. Many biological systems, however, do not have perfectly defined start and end points, and are simply feedback loops that continually self-regulate, as may be the case here.

Lastly, DNA methyltransferase 3-like (DNMT3l), while not a DNA methyltransferase itself, has strong homology to the DNMT3 family (Aapola et al., 2000). Initial characterization of DNMT3l has found that it can bind physically to DNMT3a and 3b (Okano et al., 1998), stimulate DNMT3a activity (Chedin et al., 2002), and help localize histone deacetylases (HDACs) (Aapola et al., 2002) .

DNA methylation in general causes a problem for DNA repair machinery, in that a particular kind of DNA damage - spontaneous deamination - causes methylcytosine to change into thymine. As thymine is another standard DNA base, there is no obvious fault for the DNA repair machinery to correct. By contrast, the spontaneous deamination of unmethylated cytosine creates uracil. As uracil is not normally a base found within DNA, repair enzymes can identify and remove it. While the rate of spontaneous deamination is fairly low (Shen et al., 1994), over the eons that CG methylation has been utilized by vertebrate evolution this undetectable change from C to T has led to an imbalance of CGs in the genome, with CGs representing only 1% of dinucleotides in the genome compared to > 4% frequency for all other dinucleotides pairs. Methylation is very prevalent among these remaining CGs, with > 70% of all CGs methylated in humans (Ziller et al., 2013).

This methylation is not evenly distributed, and neither are the underlying CGs. Studies examining CG density and methylation across the entire human genome have found that many CGs are found clustered together in what are known as “CpG islands.” Many of these clusters are found in transposon repetitive elements, and are at least as highly methylated as the rest of the genome. The remainder (~40,000 out of ~350,000 total, depending on the exact definition used) are in the non-repetitive genome, are generally un-methylated, and often (~50%) overlap with transcription start sites (TSS) and gene promoters (Glass et al., 2007). The remaining 50% that do not overlap with TSSs are at

genomic locations often display signs of transcription (Maunakea et al., 2010), making this genomic feature a very common de facto promoter element.

This relationship (un-methylated CpG islands at the promoters of genes) has given rise to the classical inverse relationship between DNA methylation and gene expression, where lower "levels" of DNA methylation in a particular genomic region are associated with increased gene expression. There are two main biological mechanisms for this association. The first is a direct interference of the ability of many transcriptional activators to bind to methylated DNA, usually through steric interference; the second is the ability of methylated DNA to recruit inhibitory factors (often through the mediation of methyl-binding protein 1 and 2 (MeCP1 and MeCP2) that in turn will interact with other proteins to spread inhibitory histone marks and a condensed chromatin structure (Tate and Bird, 1993). These mechanisms and inverse association, however, are not as straightforward outside CpG islands.

The several thousand bases outside these CpG islands have in recent years become known as "CpG island shores" or "CG shores." These regions are also relevant epigenetically, as most DNA methylation differences between tissues are found in shores (Irizarry et al., 2009). The relationship of methylation to gene expression in these regions is much more complicated than within CpG islands and appears to depend on the precise location of methylation within the genome and the identity and function of other factors binding in that region. In addition to its classical role in regulating expression and promoter CpG islands, methylation is known to promote and repress alternate transcription start sites (Maunakea et al., 2010), affect mRNA splicing (Oberdoerffer, 2012), modulate splicing sites within genes (Maunakea et al., 2013), positively affect gene expression when within gene bodies (Yang et al., 2014), and promote gene expression when at repressor binding sites (Ando et al., 2000). Generally, the biological mechanisms for these processes appear to be the same – direct interference with TF binding or the creation of binding sites for proteins like the MeCPs. The difference appears lies in location; it is easy to see how interference with the binding of an expression-promoting factor at a gene promoter will

have opposite effects as interference with the binding of a transcriptional repressor at an upstream enhancer element.

Within the cell's nucleus, DNA is not simply a tangled bundle of double helix spaghetti. DNA is almost all wound and packaged around proteins known as histone complexes. Each histone complex consists of eight subunits: two each of the histone proteins H2A, H2B, H3 and H4. The unit of a histone complex and the approximately 147 bases of DNA wrapped around it is called a nucleosome, and these nucleosomes comprise the fundamental unit of chromatin – a term for the overall organization of DNA and associated macromolecules within a cell. The most obvious purpose of this wrapping is packing. If strung out, the DNA in each human cell would be almost five feet long. Wrapped around histones, however, the nucleosomes only stretch out 90 micrometers (Redon et al., 2002).

Histones, however, are not merely spools around which DNA is wound. The H3 and H4 subunits have "tails" that protrude, and these tails can greatly influence the types of proteins that bind to DNA. An enormous variety of post-translational covalent modifications occur on these tails, creating a diverse mosaic of possible binding sites for other proteins. Taken together, the sum of these post-translational modifications are known as the "histone code" and comprise at least six different types of post-translational modifications occurring at different amino acids along the H3 and H4 tails. As each combination of different histone marks at different places potentially comprises a binding site for a different protein, the histone code has been postulated to be a cellular "language" with a vocabulary consisting of thousands to millions of words.

A full review of every histone modification and position is beyond the scope of this work, but one example will be presented here. Histone acetylation has been observed to occur on at least five different lysines on the H3 tail and four lysines on the H4 tail (Cell_Signaling_Technology, 2014). These marks are placed by a class of enzymes known as histone acetyltransferases (HATs), of which there are at least five different families, and removed by enzymes known as histone deacetylases (HDACs), of

which there are at least four major types. Normally, lysines have a positive charge. Histone molecules are fairly lysine rich, and it is thought that the interaction of the negatively charged DNA and the positively charged lysines helps stabilize the interactions between histones and DNA. Lysine acetylation, however, eliminates the positive charge of the lysine, leading to decreased interaction between DNA and histone, and ultimately in DNA that is less tightly bound. This has the effect of making the DNA more accessible to transcription factors and transcription machinery, allowing more gene expression. In addition to this physical effect, however, many transcription activating proteins have bromodomains that bind directly to acetylated lysines on the histone tails, bringing them into close contact with the DNA. The end result of this acetylation is a strong correlation between histone acetylation and gene expression, especially at gene promoters.

Epigenetics, however, is not merely the modulation of gene expression; it is also the perpetuation of these modulations through cell division. Unlike DNA methylation and the DNMT1 enzyme that provides a clear and easy mechanism for its perpetuation through DNA replication, no certain mechanism of histone code cellular memory is known. Without such a mechanism, any changes that occur to histone marks must necessarily be transient and would probably not make a lasting difference to the organism. This is somewhat difficult to square with the known specificity of histone modifications. Histone acetylation, for example, does not occur randomly throughout the genome, but instead occurs specifically at active genes, and at the same active genes in daughter cells. Is there a system to simply place histone marks anew in the correct places at the end of every cell division? Or are there systems in which the actual location of histone marks are copied over onto newly synthesized DNA through mitosis?

Recently, several theories have been proposed as to a mechanism for histone code cellular memory. As an example, EZH2 is the enzyme that causes tri-methylation at lysine 27 on H3. It has been observed that EZH2 not only methylates this lysine, but continues to co-localize with this mark even

during DNA replication (Hansen et al., 2008). It has therefore been proposed that this close association with H3K27me3 domains could allow the methylation of the newly synthesized DNA almost as soon as it is made, essentially copying the information from the old strand to the new. If true, this would represent a form of memory through cell division that would elevate at least one histone mark to full epigenetic status.

Thus far, only the normal functioning of epigenetics has been reviewed – what happens if these systems become disrupted? Because of the persistence of epigenetic marks, disruptions and deleterious changes could be maintained by cells and their descendants, slowly spreading throughout the organism, much like a genetic mutation. Unlike a genetic mutation, however, epigenetic disruptions could be acquired long after birth by anyone, regardless of the genotypes of their parents. Indeed, a growing body of literature, including work discussed in the later chapters of this thesis, show that deleterious disruptions in the systems of epigenetic mark placement and maintenance (epi-mutations) can be found in many long-lasting conditions and diseases and can be initiated by common environmental factors affecting everyone.

The first major human disease found to be systematically associated with wide-spread epigenetic changes was cancer (Feinberg and Vogelstein, 1983). Exact patterns of differential methylation in cancers have been shown to be idiosyncratic both for different types of cancers and even for different individual tumors of the same cancer type. Despite this, however, there are several overall patterns observed in the majority of cancers – large-scale loss of DNA methylation (hypomethylation) and region-specific gain of DNA methylation (hypermethylation). As an example, a recent study of colon cancer found many differentially methylated regions (DMRs), with most of the hypermethylation taking place near CpG islands, and most of the hypomethylation taking place further away, at the CpG island shores that normally have colon-specific methylation patterns (Irizarry et al., 2009).

Interestingly, despite large epigenetic changes taking place in nearly all examined cancer types, it has been difficult to come up with easily usable DNA methylation biomarkers for cancer. This appears to be due to the overall variability of individual marks – while there are changes taking place in most cancers, the odds of any individual CG being differentially methylated in a specific tumor is less than definitive. This variability, however, may be in itself a biomarker. Recent work examining specific genetic regions in cancer versus normal samples has found that while most regions feature some individual tumors having the same methylation levels as normal tissue, rendering them difficult to use as a biomarker, the examination of multiple regions across the genome allowed tumor samples to be identified as having overall increased methylation variability (Hansen et al., 2011). This increase in variability even appeared to progress with tumor metastasis, making the measurement of methylation variability a potential assay for the future metastasis potential of apparently benign tumors. The general biological hypothesis explaining how increased epigenetic variability in itself could lead to more pathogenic tumors is based in selection (Ohlsson et al., 2003). While not all cells will have epigenetic changes that promote the unrestricted growth common in cancer (increased expression of growth genes, decreased expression of apoptotic or tumor-suppressor genes), in a large population of cells, statistically at least a portion of the cell population will have these types of changes, and these cells will begin to propagate without restraint.

Dysregulated epigenomes are not limited to cancer, however. Theoretically, epigenetic changes provides an excellent candidate mechanism for any disease that is persistent but not genetically determinative. In the last decade, investigators have found epigenetic changes in a wide of range of different diseases and conditions. Aside from cancer, epigenetic changes have been associated with systemic lupus erythematosus (Javierre et al., 2010), autism (Nguyen et al., 2010), Type 1 Diabetes (Rakyan et al., 2011), Type 2 Diabetes (T2D) (Volkmar et al., 2012) (Multhaup et al., 2015), coronary

heart disease (CHD) (Sharma et al., 2014), rheumatoid arthritis (Liu et al., 2013b),and bacterial inflammatory responses (Shuto et al., 2006).

Interestingly, many of these diseases are known to have a genetic component as well. Both T2D and CHD have over 50 individual genetic single nucleotide polymorphisms (SNPs) associated with risk for these diseases (Welter et al., 2014). In both these diseases, the cumulative effect of these SNPs can only explain less than 20% of the disease risk, and the genetic basis of these diseases as a whole can only explain ~50% of disease risk (more detail on these two diseases in Chapters 2 and 4 of this work). This makes sense, as environmental factors, such as diet for T2D and smoking for CHD, are known to play huge roles in the development of these conditions (Kannel et al., 1987).

Epigenetics is known to be influenced by both the environment and by the genetic landscape. Recent studies have shown large epigenetic changes between genetically identical animals that have received different diets (Multhaup et al., 2015). On the genetic side, work on blood from rheumatoid arthritis patients have shown that particular SNPs are tightly associated with the methylation status of individual CpGs (Liu et al., 2013b), implying that the genetic makeup of an individual can also affect the epigenome. Together, these two findings show the potential for epigenetics to actually mediate between the genetic and environmental bases of many common human diseases, a potential that has been invoked as a hypothesis for the basis behind the post-childhood onsets and progressive nature of many common human diseases (Bjornsson et al., 2004).

The methods used in the study of epigenetics have evolved over the years. The original methods were chromatography based and could only be used on bulk DNA (Hotchkiss, 1948). DNA methylation analysis was transformed with the advent of two technologies – bisulfite treatment and sequencing. The chemical bisulfite (hydrogen sulfite), when added to DNA, will convert all cytosines into uracil by bringing about deamination as discussed above. After amplification, usually with polymerase chain reaction (PCR), these uracils are replaced with thymines, resulting in the conversion of all C nucleotides

into T. Methylated DNA, however, is protected from this process. The combination of bisulfite and PCR allows the measurement of methylation (by examining the placement of the only remaining Cs in the genome) after amplification. This measurement is usually accomplished via sequencing. Originally, Sanger sequencing was used to measure DNA methylation in a technique called cloning bisulfite sequencing (Zhang et al., 2009). In this method, DNA was extracted from a sample of interest, bisulfite treated, and then the genomic location of interest was PCR-amplified. These sequences were then cloned into bacteria such that bacterial colonies represented individual DNA strands from the original sample. The sequencing of a sufficient number of these bacterial colonies would then give information on the same number of individual strands from the original sample, thus allowing quantification of DNA methylation.

This approach had a number of drawbacks, however, especially involving the effort and intermediate step of moving through bacteria and also the inability to examine more than a few regions at a time. The problem with having to use bacteria as an intermediary step, however, was solved with the creation of the bisulfite pyrosequencing method (Tost and Gut, 2007). In this method, DNA is extracted from the sample of interest, bisulfite-treated, PCR-amplified, and then sequenced directly by pyrosequencing. Pyrosequencing is a method in which SNPs can be quantified with a high degree of precision (within 5%) via light released by a reaction involving a series of enzymes during DNA polymerase procession over the site of interest. The increased specificity in determining SNPs allows the quantification of the percent cytosine at a particular genomic location, and thus the quantification of the amount of DNA methylation originally at that location.

The lack of throughput associated with only being able to examine a specific area of the genome were ameliorated with microarray technology. Briefly, microarrays used to examine DNA have millions of single-stranded DNA fragments attached with microscopic precision to a small slide. When sample DNA is made single stranded and allowed to wash over this slide, it will bond to corresponding DNA

fragments on the microarray if present. The amount of DNA bound to the slide (corresponding to the amount of DNA in the actual sample) can then by quantified via the attachment of fluorescent dyes to the ends of these fragments and then by taking an incredibly high-resolution picture of the resulting pattern of light. With the knowledge of the original location on the slide of the attached DNA, the amount of light corresponding to each original DNA fragment can then be assigned (by computers) to a unique position within the genome, allowing quantification of the amount of DNA from that position that was in the sample.

This technology was paired with chromatin immunoprecipitation (ChIP) and DNA immunoprecipitation (DIP) to examine DNA methylation and histone marks genome-wide. In these techniques, antibodies against either specific histone marks or DNA methylation are used to create samples enriched for the antibody ligand. These enriched samples are then paired with non-enriched samples, and labeled with different fluorescent dyes on the microarray. By comparing the amount of the dyes at a particular spot on the microarray, the relative enrichment of the immunoprecipitated mark of interest at the corresponding genomic location (and all genomic locations on the array) can be ascertained. Other methods can also be used to create enriched samples for microarraying as well. For instance, one method used in this thesis, comprehensive high-throughput array-based relative methylation (CHARM), uses an enzyme pair known as McrBC that cuts DNA only at methylated cytosines (Irizarry et al., 2008). Putting two differentially labeled samples on the microarray – an untreated sample and a DNA sample cut with McrBC – allows the quantification of relative amounts of DNA methylation at specific genomic loci.

While originally small, containing only several thousand DNA segments, microarrays are now very complex, with the popular Illumina Infinium BeadChip 450k array being able to measure more than 450,000 individual sites. Other, more expensive arrays, such as the latest version of CHARM, can

measure millions of individual genomic sites (Multhaup et al., 2015). This method of genome-wide epigenetic measurement, however, is rapidly being superseded by next-generation sequencing.

Next-generation sequencing is a term for a small collection of different technologies that can sequence large amounts of DNA very quickly and relatively cheaply. In bisulfite sequencing, DNA from a sample is bisulfite treated, "libraries" are created from it (the DNA is fragmented and modified with special ends), and then fed directly into the next-generation sequencing machine. One example, the Illumina Hiseq, then uses a method called sequencing by synthesis to determine the makeup of the DNA sample. In this method, single-stranded fragments of DNA are equipped with a primer, and then the corresponding strand is synthesized. The synthesis is done with fluorescently tagged nucleotides, with the fluorescent specific for each nucleotide. After each nucleotide is added, a laser excites the tag such that it produces light, allowing the determination of exactly which of the four nucleotides attached to the DNA. As next generation systems are capable of probing millions of sites at the same time and rapidly progressing through the synthesis reactions, this will result in obtaining data on huge swathes of the genome, more than allowable with microarray technologies. While the cost of this technique is still higher than microarrays, it is rapidly decreasing, and within a few years next-generation sequencing will become the standard.

Once data is generated by any method, it must be analyzed, and large datasets present their own analysis problems. The scientific community has responded with a large variety of analytic solutions, some of which are used in this thesis and will be discussed here.

The first problem facing analysis is normalization. Most of the methods discussed above involve comparing the amounts of multi-colored light. This has to be translated into numbers suitable for analysis. This is not a simple procedure. As an example, consider readouts from two probes from the CHARM array discussed above. One probe has half as much light associated with the methyl-depleted sample as the other, and 75% as much light associated with the untreated sample. How much more

methylation does the first sample have? Please also consider that different sequences in the underlying DNA will have effects on the intensity of signal (Aryee et al., 2011).

In this work, this problem is solved through the use of subset-quantile normalization both for CHARM (Aryee et al., 2011) and for analysis of Illumina Infinium 450k BeadChip arrays (Aryee et al., 2014). This method takes control probes – probes designed specifically to have no methylation – and uses these as "anchors" to create a known bottom for the curve of methylation values. The values for the other probes on the array are then matched to a methylation curve anchored by these control probes and scaled such that the probes with the highest values represent 100% methylation. This is done within each individual sample. This procedure is also carried out between samples, with corresponding control probes being matched across samples to normalize distributions and then quantile normalization occurring, with the distributions of non-control probes being matched to the corresponding probes on the other samples.

Once normalization is completed, the data now represents a set of reliable values. Each sample, however, may not be as reliable. Quality control (QC) is needed to identify samples that may have less high-quality data than the others. There are many reasons for technical sample failure. Individual samples may come from higher- or lower-quality DNA, and this can be reflected in the final data. Samples can be switched during processing and this is unfortunately common in studies with large numbers of samples (Lynch et al., 2012). Errors can be made during the processing of individual samples, causing them to fail to amplify as well as others or to hybridize to the DNA fragments on the microarray. All of these effects must be looked for and affected samples removed. Most of the techniques used to identify these samples are the same as used to look for batch effects.

In addition to potential severe problems with individual samples, studies with large numbers of samples can suffer from issues that can affect large numbers of samples in more subtle ways – "batch effects." These are issues that will generally affect groups of related samples and cause the measured

methylation values to be not noticeably wrong, but certainly different from other groups of samples, potentially affecting any comparison that involves samples from each group. Generally batch effects will arise between samples that were processed at different times, but the source of these effects can be due to a variety of factors. Protocols can be done slightly differently on different occasions, often in innocent ways – possibly slightly longer washes or sample thawing times – and these can have small but noticeable effects. Large microarray batches have been attributed to different technicians following the same protocol (Leek et al., 2010). Time of day or year that the samples were processed can create batches, possibly due to different amounts of ambient ozone in the atmosphere slightly changing the fluorescence chemistry (Fare et al., 2003). The list of specific causes is nearly endless, and many simply cannot be fully controlled for.

Luckily, however, there are methods to identify and compensate for batch effects. The first method that should always be employed, even before starting the experiment, is stratification. If the experimental question being asked – e.g., comparing fat to lean mice – is completely confounded with a batch effect – e.g., all fat mice are in one batch and all lean mice in another – then there will be no way to separate out the effects of the scientific question and the batch. Therefore, samples from different experimental groups should be separated out – stratified – across foreseeable batches. For microarray experiments, this means separating out case and control samples across different plates of samples and different days of sample processing. In contrast, if there are individual samples that are tightly linked – e.g., samples that are paired across a timepoint – then these samples should be processed together in order to reduce variability.

Once an experiment is performed, the first thing that should be examined is the absolute intensity of the results from each samples. Samples that fail dramatically often have vastly lower intensities than successful samples, can easily be identified by looking for outliers in intensity plots, and should be discarded. Second, the overall curve of results from each sample should be examined. In

general, these are called density plots. DNA methylation in most tissues has bimodal overall values, with most DNA methylation being around 10% or 90%. If a sample, or group of samples, has a hugely different overall methylation landscape, this is also often an indication that a technical artifact has affected that sample. At this stage, however, groups of samples affected in the same way should be checked for their relation to the experimental phenotype before being removed. DNA methylation on the X and Y chromosomes can be used to identify sample sex with high accuracy. If this predicted sex does not match the annotated sex, this can be a strong indication that a sample swap has occurred, and these samples should be removed unless the sample swap can be reverse-engineered and the correct annotation assigned.

Principal component analysis (or surrogate variable analysis – almost identical mathematical operations) deserves a special mention as the best way to identify subtle batch effects. In these operations, dimensions of the greatest variability in the data are progressively extracted without regard to the original dimensionality of the data. In practice, each principal component is a number given to each sample. There are a number of principal components equal to the original dimensions of the data – e.g., for a data set consisting of only two dimensions (a plot of height versus weight), there will be two principal components. These principal components are ranked, such that the first principal component represents the largest single dimension of variability in the data. While these principal components do not correspond 1:1 with any of the original datapoints, they are often informative in and of themselves, and can usually be plotted against the original data dimensions to understand the causes of the variability they represent. As an example, consider the first principal component for a hundred samples of methylation data from fat and lean mice. Each of these samples has a number representing its contribution to the first principal component. These numbers are not evenly distributed, however. Half the samples have numbers around 1, and the other half around 2. Examining these samples, this binary distribution does not correspond to the fat or lean samples of the mice. After examining the association

of several other variables, however, it is observed that this binary distribution corresponds very well with the two different groups in which these samples were processed prior to microarray hybridization. This means that the source of the greatest variability within this dataset is not due to the primary scientific question of the mice being either fat or lean, but instead is due to a technical artifact. This situation is very common in microarray analyses and necessitates that this batch be dealt with before any comparison between experimental groups.

Once principal components are identified, what can be done about them? One possibility is that they can be removed from the data. Using either linear regression or more complicated Bayesian methods such as ComBat (Johnson et al., 2007), the effects of these batches on the data itself can be predicted and removed. This is not a perfect guarantee of removal, as these methods assume linearity of effects; if the batch has a non-linear effect on the data, or interaction effects with other phenotypes or batches, these will not be removed.

A second possibility for dealing with batch effects involves building them into the linear regression model used to predict the effects of covariates on the phenotype of interest. In general, linear regression is the most widely used method for examining the effect of one variable on another. In the context of microarray experiments, one linear model is created for each probe on the array, and this model will represent how well methylation at that probe predicts the phenotype of interest. These models can become very complicated, however, as variables other than methylation will contribute to the phenotype as well, and these should also be accounted for in the model. The most common examples include age, sex, and gender. If these are not included in the model, then differential methylation due to these variables that are correlated with the outcome of interest will often be the top results. This is usually not desired, as these results will therefore reflect differences due to these common phenotypes and not any interesting underlying biology. Any batch effects that are not previously removed should also be accounted for in the model, as well as any pertinent additional

phenotypes that do not have to do with the biology of interest – smoking in a study of heart disease, for example.

Once linear models are created for each probe, additional data analysis can be performed. For example, the CHARM results discussed in chapters 2 and 3 of this thesis are created by using the "bumphunting" algorithm (Jaffe et al., 2012). Briefly, this method uses linear models to obtain a beta coefficient – summarizing the association of methylation with the phenotype of interest – for each probe. These coefficients are examined genome-wide, and a cutoff (usually around the 95% percentile of beta-coefficient values) is created. Loess smoothing is used to create average beta-coefficients between probes representing sites close to each other in the real genome. A sliding window method is then used to identify "bumps" – groups of CpGs close together on the genome where the average beta-coefficients are above the cutoff. These represent areas of the genome where methylation is associated with the phenotype of interest across multiple nearby probes, a situation that may have more biological importance than differential methylation in individual CpGs.

Ultimately, this is an exciting time for the field of epigenetics. Every month brings new epigenetic associations with diseases and specific phenotypes, allowing us a closer glimpse at the biological dysregulation that underlies disease. The remainder of this thesis deals with two current published and in preparation studies into T2D and CHD, respectively.

Chapter 2

Epigenetics of obesity and insulin resistance

The material presented in this chapter has been published in Cell Metabolism (Multhaup et al., 2015) and is reproduced in part below.

Type 2 diabetes mellitus (T2D) is a metabolic disorder with a rapidly increasing worldwide prevalence. T2D affects 300 million adults worldwide and that number is predicted to rise to above 430 million by 2030 (Chen et al., 2012). Although T2D has a significant genetic risk component, as determined by twin (Wareham et al., 2002) and genome-wide association (McCarthy, 2010) studies, the heritability estimate is only 21% when looking across all age groups (Almgren et al., 2011). These low heritability estimates, coupled with the rapid increase in worldwide prevalence, suggests a strong role for environmental risk factors. These environmental risk factors continue to be identified and defined (Chen et al., 2012), and recent work on the efficacy of Roux-en-Y gastric bypass (RYGB) surgery as a treatment for obesity has found that this procedure can have a profound positive effect on T2D-related metabolic indicators (Mingrone et al., 2012).

Epigenetics, the study of non-DNA sequence based information that is replicated during cell division, such as DNA methylation, has been suggested as a natural integrator of genetic susceptibility and environmental exposure in common disease (Bjornsson et al., 2004). Epigenetics has also attracted considerable scientific and lay attention due to its dynamic nature across the lifespan (Feinberg et al., 2010), association with common disease (Cui et al., 2003), and reversibility under targeted therapies (Sharma et al., 2010). Additionally, most common human diseases are explained to a very limited degree by known individual common genetic variants, with ~3.4% of risk profile score explained for psychiatric disorders like schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics, 2014), and ~10.7% for T2D (Consortium et al., 2013). This combination of limited genetic causality, environmental influences and persistence over long time periods suggests a likely role for epigenetics in common human disease. However, epigenetic studies have their own limitations, including the need in most cases to use cells appropriate to the disease under study, confounding effects such as age, and the often considerable difficulty in designing replication sets, which are much easier in purely genetic studies because of the universality of the sample type (DNA from blood for SNPs or sequence). A number of

methodologies have been developed by our and other groups to adjust for cell type composition, confounding variables, and replication studies (which are typically much smaller) (Houseman et al., 2012; Liu et al., 2013b; Montano et al., 2013).

Genetic variants near the *FTO* gene have been linked to altered DNA methylation, though in this case the DNA methylation was not also associated with T2D (Bell et al., 2010). Analysis of known candidate genes in a small set of monozygotic twins discordant for T2D showed epigenetic changes at *CDKN2A* and *HNF4A*, but the study was not sufficiently powered for genome-wide agnostic discovery (Ribel-Madsen et al., 2012). Studies of pancreatic islets have found 254 methylation differences examining 5 deceased T2D patients and 11 controls (Volkmar et al., 2012) and 1,659 CpGs differentially methylated between 15 T2D and 34 non-diabetic controls (Dayeh et al., 2014). Similar changes have also been found in peripheral blood leukocytes from obese humans early after RYGB (Kirchner et al., 2014). RYGB is highly effective at restoring normal metabolic phenotypes in severely obese people (Kashyap et al., 2013). Moreover, interventions improving metabolic dysfunction in T2D and obesity influence DNA methylation when affected tissue is examined directly. One study that examined methylation in skeletal muscle from obese and lean subjects at 14 individually-selected loci found that methylation in obese subjects reverted to lean methylation levels after RYGB (Barres et al., 2013).

Additionally, after RYGB, there is a greatly varied phenotypic response even within usually tightly correlated variables. For example, in a study of 66 obese and T2D who underwent RYGB with a follow-up of six years, glycated hemoglobin (HbA1c) and fasting glucose levels improved rapidly for six months after surgery and then leveled off, whereas total body weight loss and lipid indicators dropped but took 48 months to reach a nadir (Cohen et al., 2012). As methylation is a relatively stable epigenetic mark, differential rates of change in DNA methylation governing pertinent genes could help to explain this response to RYGB.

In the work presented in this thesis, myself and many others (hereafter referred to as "we", please see Acknowledgements) have taken a novel approach in examining two species to identify candidate genes involved in obesity and T2D likely through epigenetic mechanisms. We first examined the epigenetic consequences of a high-fat diet in a carefully controlled experimental setting whereby obesity, due to excessive calorie consumption in the mouse, provides a close model for the insulin resistance and metabolic phenotypes that develop in human obesity and typically precede T2D. We then replicated across species—in humans—by analyzing adipose tissue from a cohort that both reproduces and reverses a phenotype similar to the obese mouse, i.e., severely obese versus lean subjects, as well as the same obese subjects pre- and post-RYGB. The use of samples from the same subjects pre- and post-RYGB allows a human isogenic comparison of the effect of obesity-induced metabolic disturbances. This cross-species approach exploits the power of evolutionary selection, whose mechanisms have survived the 50 million year separation between mouse and human, in a more comprehensive manner than simple replication from human set to human set, and may better identify functionally important environmental targets. We lastly stratified these cross-species obesity-associated regions using genetic association data from a large genome-wide association study (GWAS) for T2D to more directly link our obesity-derived phenotypes with human T2D. As a result of this approach, we are able to identify five genes with novel roles in insulin resistance, suggesting that this cross-species approach provides a powerful experimental system for identifying the genomic variation associated with common disease.

Our approach moves from directly manipulable mouse models of insulin resistance and obesity to the more limited availability of clinical material from obese versus lean humans, with GWAS data serving as a final connection to T2D genetic liability. This is the opposite direction of much GWAS-centered research, so we are describing the design in some detail here (Figure 1). To identify regions where DNA methylation levels were significantly associated with metabolic dysfunction in mice, we performed comprehensive genome-scale DNA methylation analysis on purified adipocytes from 12

C57BL/6 mice exposed to a high-fat diet and 12 C57BL/6 mice on a control low-fat diet. This diet-induced obesity model induces many characteristics of human metabolic syndrome in mice, including obesity, insulin resistance hyperinsulinemia, hyperglycemia and hypertension (Surwit et al., 1988), purely as the result of an environmental change in diet, rather than a specific mutation. This model has gained wide acceptance in the metabolic field and is used to study not only blood chemistry, but distinct changes within and between multiple tissues controlling whole body glucose and energy homeostasis (Almind and Kahn, 2004; Surwit et al., 1995; West et al., 1995). For comprehensive genomic analysis, we used the Comprehensive High-throughput Array-based Relative Methylation (CHARM) method that in its current form can assay over 5 million CpG sites in mouse and 7.5 million CpG sites in human. In contrast, the very commonly used Illumina HumanMethylation450 microarray (Sandoval et al., 2011) assays 485,000 CpGs and does not exist for the mouse. Methylation scores derived from array probes were clustered by genomic location and tested for significant phenotype association using a "bump hunting" algorithm we developed previously that models measurement error, removes batch effects and returns statistical measures of uncertainty for differentially methylated regions (Jaffe et al., 2012).

We then assessed these DMRs in a cohort of obese and insulin resistant humans using adipose tissue from 11 obese and 8 lean patients pre-, as well as 8 obese patients post-RYGB surgery. Obviously, we can only consider genomic areas that map from mouse to human, yet over 85% of the regions showing differential methylation in mouse map to the human genome, suggesting the general importance of these genes in regulating metabolism affected by obesity. The results of this comparison could identify genes important in the pathogenesis of obesity and associated insulin resistance, but they also could identify genes whose sequence confers metabolic risk depending on environmental exposure. In order to identify the latter set, we then looked for genes showing significant enrichment for T2D association using the DIAGRAM Type 2 Diabetes Genome-wide Association meta-analysis (Morris et al., 2012).

We also identified DMRs distinguishing other tissues—hepatocytes, pancreatic islets, skeletal muscle, and hypothalamus—in mice exposed to high-fat vs. low-fat diet, and utilized other subjects and datasets to perform similar analyses in human pancreatic islets and hepatocytes.

Across all five tissues examined—adipocytes, hepatocytes, skeletal muscle, hypothalamus and pancreatic islet—the strongest differences in DNA methylation associated with obesity were identified in adipocytes, with 232 or 448 differentially methylated regions (DMRs) depending on the commonly used Q-value cutoffs (controlling the false discovery rate, FDR, analogous to p-values corrected for genome-wide comparisons) of 0.05 and 0.10 (Table 1, full results in Table 2). As an example, in adipocytes from high-fat-fed mice, we found hypermethylation overlying the promoter of phosphoenolpyruvate carboxykinase 1 (*Pck1*, Figure 2A). PEPCK, the product of *Pck1*, catalyzes a rate-limiting step in gluconeogenesis, is essential for lipid metabolism in adipose tissue, is known to be regulated by insulin, and has been linked to lipodystrophy and obesity in mice (Beale et al., 2004). We also found hypermethylation in a region 2kb from the promoter of the ATP-binding cassette sub-family D member 2 (*Abcd2*) gene in high-fat mice (Figure 2A). This gene is involved in the import of fatty acids and fatty acyl-CoA into the peroxisome (De Marcos Lousa et al., 2013). These genes have not previously been shown to be altered epigenetically in obesity or insulin resistance.

In addition to the high-fat versus low-fat analysis, even more DMRs were detected when analyzing methylation differences related to the metabolic phenotypes of body weight, fasting glucose, and insulin and glucose tolerance test area-under-curve (ITT/GTT AUC) values (Table 2). One example of a mouse GTT-associated DMR is in the *Fasn* gene, which produces fatty acid synthase and is known to be involved in the development of obesity (Figure 2B) (Funai et al., 2013). Similarly, DNA methylation levels are highly correlated with body weight within the *Nbea* gene, which has been previously associated with body weight and feeding behavior (Figure 2B) (Olszewski et al., 2012). Most DMRs found were

significantly associated with more than one trait, which is not entirely unexpected as the phenotypes themselves are highly correlated (Figure 3) (Bando et al., 2001).

We additionally examined DNA methylation in pancreatic islets purified from whole mouse pancreata, purified hepatocytes, skeletal muscle and hypothalamus in the lean and obese mouse models. While overall methylation changes were strongest in adipocytes, some genome-wide significant DMRs were also found in pancreatic islets and hepatocytes (Table 1, Table 2). A strong correlation was found between mouse body weight and DNA methylation in pancreatic islets. Figure 2C depicts two regions inside the *Kcnj11* and *Abcc8* genes where pancreatic islet methylation is positively correlated with body weight. These genes are known to be involved in insulin secretion in pancreatic beta-cells, and coding SNPs in *KCNJ11* and *ABCC8* have been strongly associated with T2D-related traits (McCarthy, 2010) and with obesity in mice (Kanezaki et al., 2004). Together, these two genes jointly encode the β-cell K_{ATP} potassium channel that mediates insulin release in pancreatic beta-cells (Flanagan et al., 2009). While mutations in these genes are direct clinical risk factors for T2D, the *ABCC8* gene has not previously been shown to be epigenetically regulated in any obesity or insulin resistance model.

While individual tissues had significant DNA methylation changes (Table 1), pooling all tissues together and surveying for DNA methylation changes in common yielded no significant results. There were, however, 29 genes corresponding to 33 DMRs in common between the individual tissue adipocyte and pancreatic islet DMRs, and these are listed in Table 3.

The genome-wide significant mouse DMRs were near genes that were significantly enriched in metabolic and inflammatory pathways. We implemented gene set analyses to assess the overall biological importance of the DNA methylation changes we observed in mouse adipocytes and pancreatic islets using the Gene Ontology Enrichment Analysis and Visualization tool (GOrilla). The genome-wide significant adipocyte DMRs were near genes that were significantly overrepresented in lipid metabolic and immune/inflammatory pathways compared to the background list of genes represented on our

array, with enrichment Q-values < $9.7*10^{-3}$ (Table 4). Examining hyper- and hypo-methylated DMRs separately in high-fat-fed obese mice, we observed that the metabolic pathway enrichment was derived from genes near hypermethylated DMRs, while the inflammatory pathway enrichment was present mainly in genes near hypomethylated DMRs. Similarly, genes near pancreatic islet DMRs showed significant enrichment in secretory pathways. Examining only hyper- or hypomethylated pancreatic islet DMRs did not significantly alter the enrichment results compared with a joint analysis.

Viewed globally, these results in dietary-induced obese mice track with known patterns of pathway change in obesity and insulin resistance. Inflammatory and immune related systems are known to be upregulated in adipocytes specifically in both obesity and T2D (Gregor and Hotamisligil, 2011; Hotamisligil, 2010). These pathways, however, have not previously been shown to be significantly associated with methylation changes in a diet-induced obesity phenotype. Additionally, recent work has shown adipose *de novo* lipogenesis downregulation associated with metabolic dysfunction, possibly mediated through the action of adipose-tissue derived lipid mediators (Roberts et al., 2009). Finally, islets isolated from human pancreatic donors in late-stage T2D have impaired insulin secretion in response to glucose stimulation (Del Guerra et al., 2005).

We then tested for replication of the methylation results at nine DMRs in adipocytes and three DMRs in pancreatic islets in an independent set of 18 mice (Table 5). These DMRs were chosen so as to replicate both the largest and smallest genome-wide significant methylation changes, in order to determine whether even the least significant of our genome-wide results were valid. Mice used in the replication set were also reared on a high-fat diet but were separate from those used for CHARM. Adipocytes were extracted from these mice using the same procedures as for CHARM and assayed using bisulfite pyrosequencing, a sequencing based method of determining individual CpG methylation at high precision (Migheli et al., 2013). Nine regions found to be differentially methylated with CHARM were assayed by bisulfite pyrosequencing in adipocytes. Eight of these regions had at least one CpG showing

significant differential methylation in the same direction as detected by CHARM. For example, pyrosequencing was done to replicate the methylation differences found over the promoter of the *Pck1* gene (Figure 4A, left) and the *Runx1* gene (Figure 4A, right). All five CpGs examined for *Pck1* showed differential methylation mirroring that found in CHARM, with four of five at p-values < 0.05 (Figure 4A, bottom left panel). Similarly, all seven CpGs within the *Runx1*-associated DMR showed methylation changes in the direction indicated by CHARM, with five at p-values < 0.05 (Figure 4A, bottom right panel).

Although these were fractionated cells under investigation, to further ensure that the results were not due to changes in the distribution of cell types, i.e. cell-type shifts in the high-fat fed obese mice resulting from the infiltration of immune cells into adipose tissue, we used quantitative PCR to characterize the expression of multiple macrophage- and adipocyte-specific markers in our purified adipocyte samples from low-fat-fed and high-fat-fed mice. We saw no significant change in the levels of expression of the macrophage (inflammatory) markers *F4/80*, *Cd14*, or *Cd68*, and we did see the expected obesity-related within-adipocyte changes of the adipocyte markers *AdipoQ* and *Ccl2* (Table 6).

To examine whether these methylation changes between high-fat- and low-fat-fed mice involved changes in the expression of nearby genes, we used quantitative PCR (qPCR) to examine the expression of thirteen genes near genome-wide significant DMRs (Figure 4B). Of these 13 DMRs, all but two were within the gene itself, one was 25kb upstream of *Pcx*, and one was 108kb downstream of *Rgs3*. These DMRs were chosen to span our set of methylation results from the largest to the smallest change that still survived genome-wide significance testing. We used qPCR to examine mRNA from the same adipocytes and mice that were analyzed by CHARM. Of the thirteen genes examined, nine showed significant changes in mRNA expression in the classical opposite direction to methylation changes in high-fat fed mice, e.g., lower mRNA expression in high-fat-fed mice with corresponding hypermethylation (Figure 4B).

Furthermore, we assessed whether these DNA methylation changes correlated with previously published genome-wide gene expression data (Xu et al., 2003). This was not a perfect comparison, as these previous studies examined whole adipose tissue whereas we used purified adipocytes and because the previously published data were from mice on a high-fat diet for 18-22 weeks versus 12 weeks for our mice. Nevertheless, we saw a highly significant inverse correlation between diet-related methylation changes and diet-related gene expression changes (Figure 5A). This inverse correlation became more pronounced when examining only the genes near statistically significant promoter DMRs (Figure 5B). These results compare favorably to other functional analyses of discovered DMRs (Ji et al., 2010; Kim et al., 2010a). Taken together, these data show that we find robustly significant DMRs in mice that correlate with metabolic traits, that these DMRs replicate in separate animals, and that methylation at many of these regions appears to have a functional effect on gene expression.

We reasoned that many functionally relevant DMRs in mice exposed to a high-fat diet serve an important metabolic function that would be conserved across species and often susceptible to similar environmental cues. Therefore, to determine whether the methylation changes observed in mouse adipocytes could be replicated in an evolutionarily divergent cohort, we performed CHARM analysis on human subcutaneous adipose tissues from 8 lean subjects and 14 obese sex- and age-matched insulin resistant subjects (Figure 1).

We first examined the replication of mouse adipocyte DMRs in human adipose tissue from obese versus lean and pre- versus post-RYGB surgery. We performed our analysis at increasingly stringent levels of significance and calculated the probability of the degree of overlap occurring randomly. For example, we observed very strong overlap between DMRs in human obese versus lean adipose tissue and DMRs in high-fat-fed versus low-fat-fed mouse adipocytes (all $p<10^{-15}$, Figure 6, rightmost five bars), showing that there is a strong correlation between areas that are regulated by methylation in metabolic dysfunction in both mice and humans. We also examined the overlap of

differential human adipose tissue methylation with other mouse tissues (Figure 6, leftmost 20 bars). The overlaps between these regions in non-adipose tissues were an order of magnitude less significant than with mouse adipocytes. These data suggest not only that methylation changes in obesity-induced insulin resistance are significantly conserved between species, but also that this epigenetic conservation is tissue-specific.

Next, to actually replicate the mouse methylation changes in human (rather than merely assess overlap), we determined that out of a total of 625 genome-wide significant mouse adipocyte DMRs, 576 had homologous regions on the human genome (hg19), calculated via the liftOver UCSC tool (Hinrichs et al., 2006), and 497 had human CHARM probes within 5kb. This is a remarkably high fraction (86.3%), suggesting that our assay method, CHARM, is highly comprehensive, but also that the location of CpG regions is strongly conserved in evolution.

We then asked what fraction of these conserved methylation regions shows differential methylation in either of the two human comparisons. Of the 497 conserved DMRs, 249 (50.3%) showed significant differential methylation ($p<0.05$) between obese and lean people (Figure 1, Table 7). These numbers were similar when analyzing differential methylation before and after RYGB surgery, with 227 mouse DMRs also being significantly differentially methylated in humans. As a final, restrictive step in using human methylation to validate our mouse results, we determined that 170 (68%) of these regions had a consistent direction of methylation change between high-fat fed obese mice and obese humans, such that if a particular region had higher methylation in high-fat-fed mice, that region would also have higher methylation in obese humans and vice versa. This may be overly restrictive, but our rationale for this criterion was that the change in methylation direction would be the same, even though not all methylation changes are inversely related to gene expression. Indeed, previous work has shown that DMRs in blood and iPSC DNA can be in the same location yet have opposite directionality in mouse and human (Ji et al., 2010).

When more restrictive human methylation significance cutoffs are used, the percentage of regions with consistent directionality (true positive rate) rises, but the total number of retained regions drops, with 67/77 (87%) directionally consistent at human obesity P-values <0.005, and 25/25 (100%) consistent at P-values < 0.0005 (Figure 7A). All 170 directionally conserved regions were associated with the metabolic phenotypes of fasting glucose, GTT, and/or ITT in addition to mouse diet status. Furthermore, 134 of these regions had a consistent effect directionality between obesity- and RYGB surgery-related methylation (e.g. higher in obesity and pre-surgery and vice versa), and a further 105 had post-surgery methylation values that were in between lean and pre-surgery methylation values, i.e., regions where methylation in obese subjects appeared to revert towards a lean phenotype after surgery (enrichment $p=2.8x10^{-3}$).

In Figure 8, we present two examples of human regions that have significant methylation changes in adipose tissue, are in homologous regions of the genome as genome-wide significant mouse methylation changes, are directionally consistent with the mouse methylation changes, and have human post-surgery methylation levels that have moved from pre-surgery levels to be closer to the lean phenotype. These regions are over two genes; *ADRBK1* (adrenergic, beta, receptor kinase 1, Figure 8A), and *KCNA3* (potassium voltage-gated channel, shaker-related subfamily, member 3, Figure 8B).

Next, we performed a similar mouse-human comparison in pancreatic islets by integrating our mouse pancreatic islet CHARM results with previously published Illumina Infinium HumanMethylation450 BeadChip array ("Illumina 450k") (Bibikova et al., 2011) DNAm data on human pancreatic islets from T2D and control subjects (Shukla et al., 2011). Given the lower coverage on the Illumina 450k compared to the CHARM platform, only 160 of the 312 statistically significant (at q<0.05) mouse pancreatic islet DMRs had any Illumina 450k probes within 1 kb. However, these probes were far more associated with human T2D status than the rest of the probes on the array ($p = 1.18x10^{-9}$, Figure 7B) demonstrating that our mouse-derived islet DMRs for weight are enriched for potential epigenetic

alteration in human T2D. In addition, 52 of the 77 mouse DMRs (67%) with corresponding significant methylation differences in human T2D had directionally consistent methylation in the human data (odds ratio= 7.2, $p=7.2x10^{-6}$).

Similarly, though we observed only a limited number of significant hepatocyte DMRs in the mouse analysis, we assayed the regions corresponding to eight of these regions in human liver tissue using bisulfite pyrosequencing. Out of these eight regions, five of them (62.5%) had at least one corresponding human CpG with statistically significant methylation changes corresponding to body weight in humans in the same direction as the mouse DMRs corresponding with body weight in mice. Methylation in the remaining three regions assayed by pyrosequencing was not significantly associated, either positively or negatively, with body weight (Table 4).

We also assessed whether the human adipose DNA methylation changes correlated with previously published human genome-wide gene expression data from obese and lean individuals (Maunakea et al., 2010). As with our mouse data, we saw a highly significant inverse correlation between obesity-related methylation changes and obesity-related gene expression changes (Figure 5A and Figure 5B, right panel).

Ultimately, we find hundreds of large, biologically relevant changes in the methylome of adipocytes that correlate with diabetes- and insulin resistance-related phenotypes. These changes are genome-wide significant even after correction for multiple testing, correlate with gene expression, separate into relevent gene ontology pathways and, most importantly, are replicated over millions of years of evolution from mouse to human. This suggests that the dysregulation of metabolic pathways observed in obesity and diabetes are not the result of specific genomic SNPs or the dysregulation of individual genes, but are rather the epigenetically mediated hijacking of large segments of metabolic regulatory systems facing environmental factors that evolution never prepared them to handle.

Chapter 3

Interaction between genomics and epigenomics of T2D

The material presented in this chapter has been published in Cell Metabolism (Multhaup et al., 2015) and is reproduced in part below.

While we have found large-scale methylation changes in both mouse and human near biologically relevant genes, T2D is known to be a disease that is in part genetically determined. There are multiple different possibilities for how these two systems could interact. They could be separate, with the genome modulating inherent disease susceptibility and the epigenome being the instrument of environmental exposure within the cell; they could be perfectly enjoined, with changes in DNA methylation essentially being the messenger between genotype and phenotype; or there could be a mix of both approaches, with methylation reflecting independent environmental influences and mediating between those and the genetic landscape. While the only perfectly deterministic way to answer this question would involve creating genetic changes and seeing how the methylome and phenotype respond (something very difficult to do in humans!) what we can do is to observe the overlaps and correlations between these two layers of information and create hypothesis based upon how they interact.

We incorporated data from human GWAS for T2D using two complementary approaches that allow further characterization of our candidate obesity-related DMRs and associated genes. GWAS summary statistics were obtained from the DIAGRAM (DIAbetes Genetics Replication And Meta-analysis) T2D genome-wide association meta-analysis, comprising data from 12 separate GWAS studies totaling 12,171 T2D cases and 56,682 controls. These separate GWAS studies have each been corrected

for population structure differences, and the meta-analysis summary statistics (e.g. test statistics and p-values per SNP) are available for public download (diagram-consortium.org). We first directly explored the association between genes with obesity-related DMRs and genes conferring clinical genetic risk for T2D by calculating statistical enrichment of the GWAS regions overlapping our DMRs. We counted the number of GWAS signals that overlapped at least one DMR and assessed statistical significance using a permutation procedure that resampled each GWAS signal 10,000 times across the mappable genome (see Methods section) (Collado-Torres and Jaffe, 2014). This procedure was performed on all significant DMRs for adipose and pancreatic islet data and then on the subset of DMRs that were directionally consistent across species. We found marginally significant enrichment for adipose DMRs, both with and without enforcing human directional consistency, among at least marginally significant GWAS signals (GWAS p-value cutoffs starting with $p < 10^{-6}$, corresponding to enrichment p-values ranging from 0.0048 to 0.0165), and non-significant enrichment for pancreatic islet DMRs (Table 8). Given the small number of directly overlapping regions, these results are likely strongly influenced by the strength of the *TCF7L2* signal. While much of the early literature on *TCF7L2* focused on its role in pancreatic islets, there is growing evidence that extra-pancreatic effects may contribute to the T2D phenotype at this locus (Prokunina-Olsson et al., 2009) (Boj et al., 2012) (Nilsson et al., 2014).

We further examined statistical enrichment in the context of regulatory networks involving genes implicated in GWAS. Although only one gene (*TCF7L2*) with genome-wide significant linkage to T2D in DIAGRAM was present in the directionally conserved cross-species adipose methylation results, the named genes at 23 genome-wide significant GWAS signals (usually the gene nearest to the lead SNP) were directly (one-step) connected to genes near DMRs either by transcriptional control or direct protein-protein interaction (Figure 9A). This amount of interaction represents significantly more than expected by random chance ($p = 0.0206$) (Figure 10), and demonstrates how genes implicated by methylation appear to be acting in the same pathways as genes implicated by GWAS. For instance,

PRC1, a regulator of cytokinesis, is associated with T2D by a genome-wide significant DIAGRAM result, but it has no known connection to any other gene implicated by genome-wide significant DIAGRAM loci. Its transcription, however, is regulated by FOXO1, an important transcription factor in gluconeogenesis, insulin signaling and adipocyte differentiation that we find to be differentially methylated in both mouse and human obesity. FOXO1 is in turn regulated by TCF7L2, one of the strongest GWAS results. These results (Figure 9A) illustrate how our methylation results fill and expand the T2D-GWAS interaction network to affect multiple pathways and multiple participants in those pathways simultaneously. Similarly, expanding beyond one-step connections, many of the 30 regions implicated by both methylation data and GWAS have extended connections to each other and act in the same pathways (Figure 9B). Furthermore, these extended pathways include many other genes implicated by both the mouse and human methylation analyses. For example, while Akt2 and Pck1 have no direct interaction, both of them interact with App, a gene implicated by genome-wide significant mouse methylation, and both of them either signal to or are regulated by CEBPB (CCAAT/Enhancer Binding Protein (C/EBP), Beta), which has nearby obesity-related methylated changes that are conserved and directionally consistent in both mouse and human and is capable of significantly affecting mouse response to high-fat diets (Rahman et al., 2012).

Given these results, we sought to further filter our obesity-related DMRs down to the subset of genes likely associated with T2D. We hypothesize that DMRs that overlap associated marker SNPs for T2D can identify genes with epigenetic mechanisms of risk in adipose tissue and/or islets. As many of the DMRs overlapping GWAS T2D loci with low p-values implicate genes already known to be involved in T2D, obesity and related phenotypes, we therefore selected the subset of DMRs within genetic loci that had at least marginal statistical association with T2D clinical risk.

This approach reduced the 170 regions of directionally consistent and evolutionarily conserved methylation change in adipose tissue using the SNP-level summary statistics of the DIAGRAM analysis

(Figure 1). In all, 30 cross-species and directionally conserved adipose DMRs directly overlapped with 27 marker SNPs (or close proxies with linkage disequilibrium > 0.8) that had some evidence of association with T2D (at least $p < 0.01$, Table 9; see Methods). We also identified ten regions where conserved pancreatic islet DMRs overlap with DIAGRAM SNPs (Table 10).

While the large number of genetic loci with nominal GWAS p-values makes finding low-level enrichment problematic, cumulatively the effects of these common SNPs have been shown to have biological importance and to explain large amounts of phenotypic variability beyond that described by genome-wide significant SNPs (Yang et al., 2010).

In these 30 regions, not only have we connected methylation change to obesity-induced T2D phenotypes across two species, but the association with T2D-associated SNPs also provides a candidate mechanism for the methylation changes observed in human obesity and RYGB surgery. These 27 identified SNPs could potentially explain up to 2.69% of genetic T2D liability, while only one of these loci reached genome-wide significance in DIAGRAM (Morris et al., 2012). Even excluding this GWAS-positive loci (*TCF7L2*), which explains 1.12% of the variance alone, the remaining regions could explain up to 1.57% of genetic variance in T2D susceptibility. This analysis may not be completely independent of the previous characterization of liability (5.7% in Morris et al.) as genes identified here may lie in the same pathways as genes containing genome-wide significant SNPs, but it is in line with a polygenic contribution of lesser common variants to the disorder (Chatterjee et al., 2013).These data suggest that for at least some of these loci, genetic variation underlies changes in methylation that are causal for T2D risk. It is also possible that these regions are also susceptible to environmental factors that influence local methylation and that they therefore serve to integrate genetic and epigenetic effects.
Note that this filtering-based approach is independent of assessing the statistical enrichment of T2D GWAS signal, either at SNP- or gene-level, within our cross-species obesity-associated DMRs, an approach commonly used with GWAS summary statistic data (Wang et al., 2010). Instead, we are

combining three lines of evidence – epigenetic dysregulation following high fat diet in mouse, epigenetic directional consistency in humans, and some evidence for clinical risk of T2D – to identify *genes* we believe are functionally implicated in the pathogenesis of T2D specifically through epigenetic mechanisms related to obesity. This approach therefore does not diminish the potential function of genes with GWAS-positive statistical association for T2D or our DMRs that do not overlap with GWAS-associated SNPs for contributing epigenetically to obesity.

We hypothesized that one mechanism by which DNA methylation and genetic variation contribute to T2D risk may involve enhancer activity, as a recent study found an enrichment of T2D sequence variants in pancreatic islet enhancer clusters (Pasquali et al., 2014). Using publicly available human enhancer maps in 86 independent cell and tissue types (Hnisz et al., 2013), we found that a striking proportion of DMRs mapped to adipose nuclei enhancers and super-enhancers (which had the largest degree of overlap across all cell types). While the background proportion of overlap for CHARM was 17.2% for adipose enhancers and 3.8% for super enhancers, 40.6% (69 overlaps, $p = 1.58 \times 10^{-15}$) and 14.7% (25 overlaps, $p=5.72 \times 10^{-13}$) of the directionally consistent 170 regions, and 53.3% (16 overlaps, $p=5.65 \times 10^{-7}$) and 20% (6 overlaps, $p=3.24 \times 10^{-5}$) of the further 30 GWAS-associated regions above lie in adipose enhancers and super enhancers, respectively (Table 11). Thus, a major mechanism for methylation-mediated metabolic dysfunction is likely through epigenetic modification of enhancers. Note that most of these enhancers were not previously known to be related to T2D through conventional GWAS or other methods.

In order to establish that our cross-species method can identify functional genes implicated in obesity, insulin resistance, T2D, and related research, we selected six genes to further investigate. Out of the set of 30 cross-species conserved DMRs that also overlay variants conferring at least marginal risk for T2D, 14 had no previously reported independent association with T2D. Of these 14 DMRs, 13 of them had obese human methylation that reverted toward the lean methylation phenotype after RYGB

surgery. As RYGB is a targeted, environmental therapy that improves multiple deleterious phenotypes including insulin sensitivity, we hypothesized that this subset of our results would be the most likely to have an effect on T2D- and obesity-related phenotypes. We then examined the physiological effect of altering the expression of these genes on adipocyte cell culture models using insulin-stimulated glucose uptake assays. This procedure can measure the responsiveness of adipocytes to insulin, an important measure of insulin sensitivity and resistance, and reduced glucose uptake has been linked to both obesity (Virtanen et al., 2002) and T2D (Martin et al., 1992). We assayed six 3T3-L1 adipocyte cell lines. Each line was stably expressing shRNAs or stably expressing expression plasmids against one of the six selected genes. In order to mimic the effects of a high-fat diet, genes hypermethylated in high-fat adipocytes were knocked down, and genes hypomethylated were overexpressed (with the exception of *Car5a*, which was knocked down as there was no suitable plasmid available at the time experiments began). Significant changes in glucose uptake were found in five of these six 3T3-L1 adipocyte cell lines.

Three of these genes (*Mkl1*, *Plekho1*, and *Tnfaip8l2*) were hypomethylated in high-fat-fed mice and obese humans, had increased expression in at least one of the corresponding high-fat/obese gene expression cohorts, and, when overexpressed in 3T3-L1 adipocytes, decreased glucose uptake in response to insulin (Figure 11). Therefore, these three associations appear to follow the classic paradigm of methylation inversely correlating with gene expression leading to a change in phenotype. *Gstz1* also followed the classical paradigm in so far as it was hypermethylated in high-fat-fed mice and had decreased gene expression in obese humans, but 3T3-L1 adipocytes expressing shRNAs against this gene exhibited no significant change in glucose uptake.

Two other genes (*Tmcc3* and *Car5a*, which overlaps with the putative mouse transcript BC048644) had altered methylation in high-fat mice and obese humans, but had no significant corresponding gene expression changes (Figure 11B). We note that this proportion of methylation-gene expression correlation (4/6 DMRs having inverse correlation, 2/6 having no or positive correlation) is

almost exactly what is expected based on overall methylation-gene expression correlation both in this work (Figure 5) and in previously published literature (Barres et al., 2013). When the expression of *Tmcc3 and Car5a* was knocked-down in cell-culture adipocytes using shRNAs, however, there were significant changes in the insulin-stimulated glucose uptake. Presumably, methylation acts on the expression of these genes in some other manner than the classical inverse correlation paradigm, or, at least, does not alter transcription in baseline high-fat-fed mouse or obese human adipose tissue.

In mouse, we identified 625 genome-wide significant differentially methylated regions (DMRs) that correlate with diet-induced obesity phenotypes in adipocytes. Of these regions, 249 had significant conserved methylation changes in human obesity, and 170 of these had the same direction of methylation change in both species. Thirty of these DMRs also overlapped with SNPs or nearby proxies that have been associated with human T2D genetic risk. These data show for the first time that DNA methylation changes in metabolic disease are conserved across species and that this conservation overlaps genomic regions where genetic polymorphisms have been associated with T2D. Our approach combines three lines of evidence – epigenetic dysregulation following high fat diet in mouse, epigenetic directional consistency in humans, and some evidence for clinical risk of T2D – to identify *genes* likely functionally implicated in the pathogenesis of T2D specifically through epigenetic mechanisms related to obesity.

Of the 30 adipose DMRs that survive our filtering process, only two are near a gene (*TCF7L2*) that has previously been associated by conventional GWAS to T2D, though one other, *ETAA1* has suggestive but not genome-wide significant associations with waist-hip ratio (Liu et al., 2013a). In the present study, while we use nominal P-value significance to identify human methylation and GWAS results, we first perform a multiple comparison correction in our initial set of mouse DMRs using a false discovery rate algorithm. As there is a growing awareness that the cumulative effect of common SNPs with low minor-allele frequency scores potentially explain large amounts of phenotypic variability

beyond that of genome-wide significant SNPs identifiable by GWAS (Yang et al., 2010), approaches like ours that can use alternative methods to identify significant areas of potential genetic risk are necessary. The unique SNPs in these regions potentially account for 2.76% of T2D genetic variance, almost half of which is known by purely genetic analysis and may be epigenetically mediated.

We observed significant changes associated with five out of six genes assayed by insulin-stimulated glucose uptake assay, a common indicator of insulin resistance. Screens using this assay and performed on sample sets not enriched for genes in gluco-insulinemic pathways have found a far smaller percentage of genes that will alter glucose uptake (~10%), (Tang et al., 2006), indicating that our method can successfully select potential targets with a much higher than random probability of affecting insulin sensitivity. Three of the genes that we found had altered glucose uptake fell into the classical inverse methylation-gene expression correlation: *Mkl1*, *Plekho1* and *Tnfaip8l2* were all hypomethylated in high-fat-fed mice and obese humans, had increased gene expression in corresponding subjects, and, when these genes were overexpressed in cell culture adipocytes, exhibited decreased glucose uptake in response to insulin, which would fit with the increased insulin resistance commonly observed in obesity and diabetes. While none of these genes have previously published roles in insulin resistance, all have been associated with inflammation, and inflammation as a gross phenotype has been previously associated with obesity and insulin resistance (Xu et al., 2003). More specifically, Mkl1 is known to be a transcriptional coactivator of serum response factor (SRF) (Cen et al., 2003), and SRF transcriptional activity has been associated with insulin resistance in skeletal muscle (Jin et al., 2011). While *Tnfaip8l2* has not been directly linked to obesity, insulin resistance or T2D, it is upregulated in diabetic rat kidneys (Zhang et al., 2010). Finally, PLEKHO1 has recently been shown to inhibit AKT/PI3K signaling (Zhang et al., 2014), a pathway known to be involved in insulin signaling.

Two other genes, however, did not fit the classical inverse methylation-gene expression pattern. While *Tmcc3* and *Car5a* exhibited hyper- and hypomethylation, respectively, in obese subjects, and their

knockdown with lentiviral shRNAs significantly altered insulin-stimulated glucose uptake assays, there were no corresponding gene expression changes in either human or mouse. A necessary role for *Car5a*, has been implicated for the optimal function of Pcx in liver (Lynch et al., 1995), and it could play the same role in adipose tissue. *Pcx*, also differentially methylated in our analysis, is the gene for pyruvate carboxylase, which converts pyruvate to oxaloacetate at the beginning of the *de novo* lipogenesis / glyceroneogenesis pathway, and in which Pck1 subsequently converts oxaloacetate into phosphoenolpyruvate. Taken together, Car5a could work with the already established Pcx and Pck1 in the alteration of *de novo* lipogenesis in adipocytes, a process essential for normal fatty acid metabolism in adipocytes (Beale et al., 2004). The role of Tmcc3 in the pathogenesis of insulin resistance in type 2 diabetes and obesity is elusive; however, one recent poster presentation at the AACR 103rd Annual Meeting in 2012 (Wang, 2012) did report evidence that knocking down TMCC3 in breast-cancer stem cells inhibited Akt phosphorylation. If this finding replicates in adipose tissue (especially for Akt2), it could potentially explain the involvement of Tmcc3 in insulin-mediated glucose uptake.

It is worth noting that as these genes did not contain common variants that passed the genome-wide significant GWAS threshold, they would not have been identified by GWAS alone. Similarly, only three out of these five genes had significant gene expression changes. This functional assay illustrates how our method of combining cross-species methylation data with GWAS results for common SNPs can implicate genes that would not have been detected otherwise. For the two genes without significant gene expression changes, we hypothesize that DNA methylation may be affecting these two genes via one of the multiple pathways not involving blocking transcription factor binding at promoters, such as splicing (Shukla et al., 2011), the regulation of alternate promoters (Maunakea et al., 2010), or interference with repressor binding at non-promoter binding sites (Ando et al., 2000). Finally, with regards to the direction of glucose uptake change, we note that insulin signaling induces both positive and negative feedback within affected cells (Gual et al., 2005), and without a methylation-gene

expression candidate mechanism it is not possible to determine which feedback loop the methylation-implicated genes are involved with.

Collectively, our results have therapeutic implications for the treatment and prevention of metabolic disease. Of the three major classes of drugs commonly used to treat T2D, thiazolidinedione, biguanides and sulfonylurea, only thiazolidinediones act on adipose tissue. This class of drugs is known to affect the PPAR transcription factors. PPARG, in particular, is a master regulator of adipocyte function, and consequently, thiazolidinediones have been shown to affect many different areas of adipocyte function, including adipogenesis, insulin signaling, and inflammation (Arner, 2003). Thiazolidinediones, however, also produce significant deleterious edema in a subset of patients, presumably due to the targeting of PPARG in the kidney, as well as adipose tissue and the liver (Kiryluk and Isom, 2007). Antagonists against specific elements of the pathway downstream of PPARG that are known to be dysregulated in T2D and obesity, such as those we identify in this study, may prove to recapitulate many of the beneficial effects of thiazolidinediones without the non-specific side effects.

Recent work in our laboratory has identified regions of the genome where DNA methylation acts to mediate a genetic effect on rheumatoid arthritis (Liu et al., 2013b), and the methylation changes in obese humans could potentially act in an analogous role. Our results in obese and insulin-resistant mouse models, however, identify methylation differences even between inbred mice, and thus are definitively the result of environmental stimuli rather than a genetic underpinning. The fact that we see many of these same methylation changes in obese humans, and that these changes are located over regions with known genetic links to T2D, implies that DNA methylation levels could be integrating and mediating genetic and environmental causes of metabolic disease at specific genomic loci.

A previous study has investigated human adipose tissue from healthy individuals before and after a six-month exercise intervention, as well as adipose tissue from individuals with and without a family history of T2D. While no significant methylation differences were found to be associated with a

family history of T2D, exercise regimens in healthy individuals were associated to genome-wide significant methylation differences near 7,663 unique genes, and more specifically near 18 candidate genes associated with T2D by a genome-wide association study (Ronn et al., 2013). Of these 18 candidate genes, only one (*TCF7L2*) overlaps with our results, suggesting that exercise-induced methylation change in healthy individuals and severe obesity induce differential methylation in mostly separate pathways. This locus was also previously implicated epigenetically in T2D but in a study of whole blood (Toperoff et al., 2012) and also was shown to contain a SNP that modified a CpG site and thereby associated with differential methylation in this region (Ronn et al., 2013). However, this particular variant was not the causal variant attributed to the GWAS signal (Gaulton et al., 2010).

It is encouraging that many of the new genetic associations described here show pathway relationships to known genetic associations. We find that a large and significantly enriched number of the genes near directionally conserved cross-species DMRs had direct protein-protein or transcriptional connections to genes connected to T2D with genome-wide significance by the DIAGRAM GWAS meta-analysis (Figure 9A). Moreover, many of the genes implicated in this study complement and expand pathways already known to be involved in T2D and insulin resistance. Furthermore, combining genes from all levels of this study creates regulatory networks that include genes with known involvement in T2D but also incorporate closely connected genes with no previously known obesity or T2D association that are shown to be involved with obesity and insulin resistance in this story. To illustrate this point, we have assembled a diagram showing 21 genes implicated by both conserved methylation and DIAGRAM overlap (Figure 9B, red and green genes). While only two of these genes (or their products) interact directly, adding additional genes implicated by genome-wide mouse methylation and conserved cross-species methylation allows the creation of a regulatory network in which all of these genes interact with each other through some degree of separation. Some of these genes, such as *FASN* and *APP*, appear to be loci in this network, and could represent potentially important targets.

There are many approaches for, and important applications of, interrogating the association of functional and genetic elements using GWAS summary statistics (Consortium et al., 2012; Jostins et al., 2012; Nicolae et al., 2010), but our approach is unique in its leverage of carefully controlled biological systems to directly integrate cross-species functional epigenomics and clinical genetic risk. Here we purposely stratify the 170 directionally consistent cross-species DMRs into the 30 that likely are most related to T2D and/or mediate genetic risk epigenetically, at least in adipose tissue. This work, of course, does not address or diminish the many GWAS associations that are not associated with methylation changes. Additionally, it is important to note that while we do not directly address the issue of methylation causality in this study, causality is, at the least, multi-tiered. Our new functional data certainly indicates that these epigenetic changes are functionally proximate to T2D-relevant phenotypes and therefore important for discovery and for clinical translation. Please note that the systems biology literature challenges conventional notions of causality as there is both positive and negative feedback in most complex living systems (Noble, 2012).

The approach described in this study may have broad applicability to identify candidate genes that may better dissect mechanisms and potential routes of treatment in common human disorders, such as cancer and cardiovascular disease. The accessibility of a limited cohort of relevant patients with well characterized clinical materials before and after disease exposure is plausible for cross-species replication. This type of analysis can generate a reliable, functional candidate disease gene set that can be used to interrogate SNP datasets and lend additional support to specific targets that would not ordinarily pass the genome-wide correction threshold. The end result is a process that can integrate information from multiple complementary sources to identify potential targets essential for the pathogenesis of common diseases, such as obesity or T2D, that do not involve highly penetrant single genes, but rather arise from multiple defects along pathways that integrate genetic, epigenetic, and environmental cues.

Materials and methods used in Chapters 2 and 3

Mouse Sample Preparation

All animal protocols were approved by the Institutional Animal Care and Use Committee of The Johns Hopkins University School of Medicine. Male C57BL/6 mice were purchased from Charles River and housed in polycarbonate cages on a 12-h light-dark photocycle with *ad libitum* access to water and food. Mice were fed a high-fat diet (HFD; 60% kcal derived from fat, Research Diets; D12492) or the matched control low-fat diet (LFD; 10% kcal derived from fat, Research Diets; D12450B). Diet was provided for a period of 12 weeks, beginning at 4 weeks of age. At termination of the study, animals were fasted overnight and euthanized; tissues were collected, snap frozen in liquid nitrogen, and kept at −80°C until analysis.

Intraperitoneal glucose and insulin tolerance tests

Cohorts of mice were injected with glucose (1 g/kg body weight) or insulin (0.8 units/kg for LFD-fed mice, 1.2 units/kg for HFD-fed mice). Animals were fasted overnight (16 h) prior to the glucose tolerance test. For the insulin tolerance test, food was removed 2 h prior to insulin injection. Serum samples were collected by using microvette CB 300 (Sarstedt). Glucose concentrations were determined at time of blood collection with a glucometer (BD Biosciences). Insulin and glucose tolerance tests were performed when mice were between 20 and 24 weeks of age.

Mouse Hepatocyte Isolation

A protocol for primary hepatocyte isolation was adapted from previously published methods (Berry and Friend, 1969; Li et al., 2010). Briefly, mice were anesthetized and a catheter was inserted into the vena cava. The portal vein was then cut to allow liver-specific perfusion. Mice were then perfused with PBS, followed by 100ug/mL Type I Collagenase (BD Biosciences) at a rate of 5 ml/min for 10 min. The liver was then removed and dissociated by straining through a 70 μm pore nylon cell

strainer (BD Falcon). The cells were then spun down and resuspended in William's Medium E (Cellgro). Primary hepatocytes were then isolated by gradient distribution via centrifugation of the resuspension in a cold Percoll (GE healthcare) solution. Verification of primary hepatocyte purity was assessed via quantitative real-time PCR for hepatocyte-specific genes compared to markers for endothelial and immune cells. We observed >90% hepatocyte purity based on gene expression.

Mouse Primary Adipocyte Isolation

Mature adipocytes were isolated from mouse fat pads as previously described (Stahl et al., 2002). Briefly, fat pads were finely chopped using scissors. Tissue was then dissociated in 2 mg/gram tissue Type II Collagenase (Sigma) in KRH buffer. The digestion was stopped by adding 10 %FBS (Atlantic Biologicals) to the mixture and cells were filtered through 100 µm pore nylon cell strainers (BD Falcon). The cells were then separated out by transferring the upper phase of cells to a new tube and washing with 5 mL of KR Buffer. The wash and resuspension was repeated 3 times and mature adipocytes were collected. Verification of mature adipocyte purity was assessed via quantitative real-time PCR for adipose-specific genes compared to markers for endothelial and immune cells. We observed >95% adipocyte purity based on gene expression.

A third set of 9 high-fat diet mice and 9 low-fat diet mice were used for replicating CHARM results using pyrosequencing. Adipocytes were extracted and purified from these mice using the same method as used for CHARM. For the pancreatic islets, however, whole pancreases were obtained from the replication mice, stained for insulin using the Anti-Insulin + Proinsulin antibody [D3E7] (Biotin) (ab20756) (Abcam, MA, USA) kit, cryosectioned into 8µm sections, and then laser-capture microdissection was used to isolate pancreatic islets (PALM Microbeam, Carl Zeiss, NC, USA).

3T3-L1 transduction and transfection

3T3-L1 cells were transducted with Sigma Mission™ lentiviral particles as per the manufacturer's protocol. Briefly, cells were plated at 60% confluency and incubated for 18 hours in a humidified incubator. Media was removed and replaced by Opti-MEM (Invitrogen) with 8µg/ml Hexadimethrine Bromide (Sigma-Aldrich). Fifteen µl lentiviral particles were added and the plates were incubated for 18 hours in a humidified incubator. Media was then removed and replaced, and on the following day media containing 10µg/ml puromycin (Sigma Aldrich) was added and the cells were cultured in puromycin thereafter.

3T3-L1 cells were transfected with overexpression plasmids using Lipofectamine 3000 (Life Technologies) as per the manufacturer's protocol. Briefly, cells were plated at 60% confluency and incubated for 18 hours in a humidified incubator. Lipofectamine 3000 (1.5µl per well containing cells) was diluted and mixed in 50µl Opti-MEM medium (Invitrogen). At the same time, 4µg plasmid DNA was diluted in 50µl Opti-MEM with 2µ P3000 reagent and mixed. The diluted Lipofectamine and plasmid DNA were then mixed, incubated for 5 min at room temperature, and distributed onto the plated cells. After 24 hours incubation, the media was replaced with growth media. After 48 hours, 500µg/ml Geneticin Selective Antibiotic (G418 Sulfate, Life Technologies) was added, and the cells were maintained in geneticin thereafter.

Lentiviral particles used: *Tmcc3* (TRCN0000126784, Sigma Aldrich), *Gstz1* (TRCN0000103080, Sigma Aldrich), *Car5a* (TRCN0000114521, Sigma Aldrich), MISSION® TRC2 pLKO.5-puro Non-Mammalian shRNA Control Transduction Particles (Control, SHC202V, Sigma Aldrich).

Overexpression plasmids used: *Mkl1* (MC202660, Origene), *Plekho1* (MC210507, Origene), *Tnfaip8l2* (MC203559, Origene), Cloning vector PCMV6-Kan/Neo (Control, PCMV6KN, Origene).

Cell culture and glucose uptake assay

Knock-down and over-expression cell lines were maintained in Dulbecco's Modified Eagle Medium (Invitrogen) supplemented with 10% FBS (Invitrogen), 10 μg/ml puromycin and 500 μg/ml geneticin (G418) as selective antibiotics for the knock-down and overexpression lines, respectively. Two days after confluence, differentiation of the knock-down lines was induced by incubation with MDI medium (4 μg/ml insulin, 0.5mM Methylisobutylxanthine (IBMX), 1.0 μM dexamethasone) for 2 days and 4 μg/ml insulin for 5 days. Differentiation of the over-expression lines was induced with MDI medium and 1 μM rosiglitazone for 3 days and 4 μg/ml insulin for 3 days. After another 3-5 days of incubation with maintenance medium, 80%-100% differentiation was shown by lipid droplet accumulation in the cells.

Glucose uptake assays were performed on differentiated knock-down and over-expression lines. After 2 h of incubation in serum-free DMEM, they were washed twice in pre-warmed PBS and placed in HEPES-buffered saline solution (25 mM HEPES, pH 7.4, 120 mM NaCl, 5 mM KCl, 1.2 mM $MgSO_4$, 1.3 mM $CaCl_2$, 1.3 mM KH_2PO_4, and 0.5% BSA) containing 10 nM or 100 nM insulin for 20 min. Then, 0.5 μCi/well 2-deoxy-D-[3H]glucose (Moravek) was added for 5 min. The reactions were terminated by two ice-cold PBS washes. Cells were then incubated for 10 min with whole cell lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, and 10% glycerol). The lysates were transferred to scintillation vials containing Ecoscint scintillation fluid (National Diagnostics) and counted with a Beckman Coulter counter (model LS 6000SC).

Clinical Cohort

This study was approved by the Regional Ethics Committee of Stockholm. All participants provided informed oral and written consent. Clinical characteristics are shown for the obese men before and after RYGB surgery (n = 11, 8, respectively) and non-obese (normal weight) age-matched men (n = 8). Full information for human subjects can be found in Table 12.

Human Sample Surgery and Subcutaneous Adipose Tissue Biopsies

A standard laparoscopic RYGB with a 1 m Roux limb was performed. The patients were weight stable and not subjected to a preoperative weight loss period. Subcutaneous abdominal adipose biopsies (50–100 mg) were obtained from the obese and non-obese (normal weight) subjects. Biopsies were obtained at the beginning of RYGB surgery (obese subjects) or elective laparoscopic cholecystectomy (lean subjects) after the induction of general anesthesia. Only non-glucose-containing intravenous solutions were administered before the biopsy was taken during RYGB or elective cholecystectomy surgery after an overnight fast. Biopsies taken from the obese subjects 6 months after RYGB surgery were obtained under local anesthesia (5 mg/ml of lidocaine hydrochloride) in the morning after an overnight12 hour fast from the same surgical incision as the initial biopsy. Biopsy samples for DNA analysis were immediately frozen and stored in liquid nitrogen until analysis. The patients were prescribed a liquid diet for 1 month after the RYGB surgery and then solid food. All RYGB patients were prescribed multivitamin, B12, folic acid, vitamin D, and calcium supplementation once daily. Fat and liver biopsies were obtained at the beginning of RYGB surgery (obese subjects) or elective laparoscopic cholecystectomy (lean subjects) after the induction of general anesthesia.

CHARM DNA methylation analysis

Genomic DNA from all samples was purified with the MasterPure DNA purification kit (Epicentre) following the manufacturer's protocol. Genomic DNA (1.5–2 μg) was fractionated with a Hydroshear Plus (Digilab), digested with McrBC, gel-purified, labeled and hybridized to a CHARM microarray as described (Ladd-Acosta et al., 2010). The mouse CHARM 2.0 array used in the analysis now includes 2.1 million probes, which cover 5.2 million CpGs arranged into probe groups (where consecutive probes are within 300 bp of each other) that tile regions of at least moderate CpG density. The human CHARM 3.0 array now includes 4.1 million probes, which cover 7.5 million CpGs. These arrays include all annotated and non-annotated promoters and microRNA sites on top of the features

that are present in the original CHARM method. We dropped 7 human arrays with <80% of their probes above background intensities, resulting in 14 pre-surgery obese samples, 8 post-surgery obese samples, and 8 lean samples that underwent DNA methylation analysis. The design specifications are freely available on our website (rafalab.jhu.edu). We then removed sex chromosomes to improve the batch correction methods.

Subsequent pre-processing, normalization and correction for batch effects were performed as previously described (Jaffe et al., 2012). Briefly, we applied our "bump hunting" approach which involves a) performing linear regression at each probe, comparing DNA methylation levels versus a covariate of interest (e.g. high- versus low-fat diet), adjusting for surrogate variables (Leek and Storey, 2007), b) smoothing the regression coefficient for the covariate of interest across nearby probes and c) thresholding these smoothed regression coefficients across all probe groups, which forms differentially methylated regions (DMRs) representing adjacent probes with statistics above the threshold. Each DMR is summarized by its "area", or the sum of the adjacent statistics above the threshold. We used the 99.9^{th} percentile of the smoothed statistics for each respective species, tissue and trait comparisons bump hunting analysis. Statistical significance was assessed via linear model bootstrapping, retaining surrogate variables, followed by bump hunting, which approximates full permutation (e.g. permuting trait, recalculating surrogate variables, then bump hunting) using much less computational time (Jaffe et al., 2012)

Bisulfite Pyrosequencing

Individual CpGs inside DMRs found as a result of CHARM analysis were chosen for validation using the MethPrimer software (Li and Dahiya, 2002). Genomic DNA (gDNA, 200 ng) from each replication sample was bisulfite treated using the EZ DNA Methylation-Gold™ Kit (Zymo research) according to the manufacturer's protocol. Bisulfite-treated gDNA was PCR amplified using nested primers, and DNA methylation was subsequently determined by pyrosequencing with a PSQ HS96

(Biotage). Artificially methylated control standards of 0, 25, 50, 75 and 100% methylated samples were created using mixtures of purified and SssI-treated whole genome amplified (WGA) Human Genomic DNA: Male (Promega). Primer sequences used for the bisulfite pyrosequencing reactions are available in Table 13.

Quantitative PCR analysis

Validated primers for all genes were taken from PrimerBank (Wang and Seed, 2003) and synthesized by Integrated DNA Technologies (Coralville, IA, USA). Exact primer sequences are available in Table 14. RNA was extracted with Trizol reagent (Life Technologies, Carlsbad, CA, USA), cDNA was created with Quantitect Reverse Transcriptase Kit (Qiagen, Venlo, Netherlands), and quantitative-PCR was performed with Fast SYBR Green (Applied Biosystems, Foster City, CA, USA) on a 7900HT Fast Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). RNA levels were normalized to same-sample 18S RNA levels.

GO annotation

We analyzed GO annotation using the GOrilla tool (Eden et al., 2009). Genes identified from our analysis were compared to a background of all genes detectable on that array to calculate enrichment.

Whole-genome gene expression analysis

Whole genome gene expression data for mouse and human analogues of our study was found and downloaded from GEO (Barrett et al., 2013). The mouse data was already pre-processed, and the human data was pre-processed using Robust Multi-array Averaging (RMA) from the Affy R library (Gautier et al., 2004). The gene expression data was then matched against the DMRs closest to corresponding genes and the log fold change (logFC) of the gene expression was plotted against the average value of the smoothed effect estimate within the DMR, and p-values were generated using t-tests based on Pearson's product moment correlation coefficient.

Enrichment between human and mouse DMRs

The liftOver tool from the UCSC genome browser transformed the coordinates from two sets of human DMRs (obese versus lean, and pre- versus post- RYGB surgery) from the hg19 human genome to the mm9 mouse genome, as implemented in the *rtracklayer* Bioconductor package (Lawrence et al., 2009). The locations of the 249,094 probe groups on the human CHARM array were also lifted over to serve as the natural background for enrichment, of which 214,646 (86.2%) had any analogous sequence in mouse, and a further 109,234 (50.9%) were within 5kb of a mouse CHARM probe group. For each pair of DMR lists, one from the two lifted over human DMRs and another from the 25 mouse trait DMRs (Table S1), we calculated the number of DMRs at given within-specific p-value significance levels, and calculated the number that overlapped within 5kb across species. Enrichment tests were chi-squared tests based on the number of species-overlapping significant DMRs, then DMRs only significant within each species, and finally the number of lifted probe group (of the 109,234) that were not significant in either species (which creates a 2x2 table of the number significant in both species, significant in just human, significant in just mouse, and significant in neither species). This is analogous to creating a Venn diagram between significant human and mouse DMRs.

Cross-species statistical analysis

We combined significant adipocyte mouse DMRs (at FDR < 5%) across the five traits (glucose, GTT, ITT, weight, and diet) by retaining the maximal coordinates over overlapping cross-trait DMRs resulting in 625 independent DMRs associated with at least 1 trait in adipocytes in mouse. These regions were lifted over from the mouse mm9 genome build to the human hg19 genome build as implemented in the *rtracklayer* Bioconductor package (Lawrence et al., 2009), of which 576 had homologous regions in the human hg19 genome. These DMRs were annotated to the nearest human charm probe group based on the annotation within 5kb, leaving 497 DMRs. We then computed a difference and corresponding p-value in obese versus lean and then in obese humans pre- versus post RYGB surgery

using linear regression, and retained the minimum p-value, number of probes with $p < 0.05$, and the slope at the smallest p-value, within each of the 497 mapped DMRs. Directional consistency across species was higher methylation in obese compared to lean and positive association between DNA methylation levels and GTT, ITT, fasting glucose, body weight, and a high-fat diet.

DIAGRAM GWAS analysis

We next integrated GWAS results into the 497 mouse DMRs by obtaining publicly available results from the DIAGRAM meta-analysis (http://diagram-consortium.org/downloads.html; Stage 1 GWAS: Summary Statistics download) with coordinates in genome build hg18. We then generated regions of high genotypic correlation by taking all SNP rs numbers with $p < 0.01$ (n=39,081) passing them through the SNAP tool using CEU 1000 Genomes Pilot 1 data (Johnson et al., 2008), obtaining proxy SNPs with $R^2 > 0.8$ (n=167,055 unique proxies), and recording the coordinate range of the proxies for each SNP. Overlapping per-SNP risk regions were merged if overlapping (n=7,946 genotypic risk regions) and the smallest p-value across all merged SNPs represented the p-value for the genotypic risk region. These genotypic regions were lifted over to hg19 coordinates for cross-species analysis as described above. We estimated the variance in disease susceptibility based on the algorithms provided in the Methods section of Morris et al (Morris et al., 2012) and from Wray et al (Wray et al., 2010) using 1000 Genomes-derived risk allele frequencies and assuming a disease prevalence of 8% for a given collection of risk SNPs.

We assessed potential enrichment between DMRs and the GWAS results using two complementary approaches – the first approach assessed the enrichment in genome location between DMRs and the LD blocks from the GWAS. This permutation-based enrichment test is performed on two lists of genomic regions (e.g. chr:start-end) that assesses the degree of overlap relative to the background genome. At a given GWAS p-value cutoff, we counted the proportion of GWAS signals that overlapped at least 1 DMR, and then generated background overlap by resampling the same number of

GWAS regions (and the same length distribution) 10,000 times from the mappable genome (e.g. the genome after removing coordinates corresponding to telomeres, centromeres and other gaps present in genome build hg19, available from UCSC). Empirical p-values for enrichment were calculated by counting the number of null proportions that were greater than the observed proportion. R code is available on GitHub (Collado-Torres and Jaffe, 2014).

The second approach assessed enrichment in gene symbols based on all genes directly connected (one-step) to genes linked to T2D with genome-wide significance by the DIAGRAM meta-analysis based on regulatory networks generated using QIAGEN's Ingenuity IPA (Ingenuity® Systems, www.ingenuity.com). These sets (also known as interaction networks in Ingenuity) were able to be generated for 57 out of 59 genome-wide significant genes. Full interaction networks were not able to be retrieved for the remaining two genes, and these were excluded from the analysis. These interaction networks then had chemicals, groups, complexes and miRNAs filtered in order to limit the potential interacting partners to genes and protein products.

We computed whether genes overlapping obesity-related DMRs were more likely to be associated with GWAS genes and their interaction networks. We first removed DMRs that were not within 10kb of a RefSeq gene, leaving 244 and 471 obesity-related DMRs in islet and adipose tissue respectively (from 312 and 576). Then we counted the number of GWAS-associated genes and their directly connected partners in the genes containing DMRs. This procedure was also performed after the cross-species conservation filtering step described above, leaving 44 and 146 conserved obesity-related DMRs overlapping genes. We obtained statistical significance based on a resampling analysis, where we resampled the same number of probes groups 100,000 times from all probes groups mapped to human genes on the mouse CHARM design by: 1) lifting the range of the coordinates of each probe group to hg19, 2) removing poorly lifted probes groups defined as greater than 1.5 times the longest (in bp) original probe group prior to lifting over, 3) assigning the nearest human gene to each lifted probe

group, and 4) dropping lifted probes groups not within 10kb of a human RefSeq gene. We counted the number of GWAS signals or their directly connected partners that overlapped the resampled genes in each iteration, and calculated an empirical p-value based on this null distribution. This procedure was therefore performed four times, for both adipose and islet DMRs with and without filtering for cross-species conservation.

Chapter 4

Epigenetic changes in Coronary Heart Disease in the general population

Coronary artery disease (CAD), also known as coronary heart disease (CHD) and ischemic heart disease (IHD), is the leading cause of death worldwide, representing 13.3% of all global mortality (Lozano et al., 2012). The lifetime risk of developing CHD has been estimated at one in two for men and one in three for women (Lloyd-Jones et al., 1999). While mortality due to CHD peaked in Western countries in the 1960s and 1970s (Cooper et al., 2000), CHD continues to spread in the developing world and in 2010, three fourths of global death due to CHD occurred in global and developing countries (Gaziano et al., 2010).

Twin studies have estimated that 30-60% of CHD risk variation can be attributed to genetics (Marenberg et al., 1994). Currently, genome-wide association studies (GWAS) have identified over 50 common genetic variants associated with coronary heart disease risk (http://www.genome.gov/gwasstudies). Cumulatively, however, these loci only explain ~10% of CHD additive genetic variance (Schunkert et al., 2011). This low proportion of explained genetic risk is in line with other complex chronic diseases such as Type 2 Diabetes.

The risk factor profile of CHD is complex, and environmental factors shown to contribute to CHD involve smoking, alcohol use, socioeconomic status, education and diet (Pearson et al., 2002) (Matthews et al., 1989). Similarly, biomarkers including diabetes status, hypertension, blood pressure, and hyperlipidemia have all also been associated with CHD (Khot et al., 2003).

Epigenetics provides a possible mechanism to explain both the low effect size of GWAS risk variants as well as how genetics and environment can interact to produce disease phenotypes (Bjornsson et al., 2004). While there have been several studies that established links between global

DNA methylation levels and CHD or CVD (Kim et al., 2010b; Sharma et al., 2008; Wei et al., 2014), there has yet been very little research examining the links between methylation at specific genomic positions and CHD.

Here we characterize whole blood genome-wide DNA methylation in 336 samples from an age- and gender-matched CHD case/control subset of the Framingham Heart Study using the Illumina Infinium 450k BeadChip array. Analyzing this paired case-control CAD cohort, we find over 20,000 significant ($p<0.05$ CpGs).

Chapter 4 Methods

The Framingham Heart Study sample

The Framingham Heart Study (FHS) is a community based longitudinal study of participants living in and near Framingham, MA, at the start of the study in 1948 (Dawber et al., 1951). The Offspring cohort was comprised of the children and spouses of the original FHS participants, as described previously (Kannel et al., 1979). Enrollment for the Offspring cohort began in 1971 (n=5,124), and in-person evaluations occurred approximately every 4 to 8 years thereafter. The current analysis was limited to participants from the Offspring cohort who participated in the 8th examination cycle (2005-2008) and consented to genetics research. DNA methylation data of peripheral blood samples collected at the 8th examination cycle were available in 2,741 participants.

All participants provided written informed consent at the time of each examination visit. The study protocol was approved by the Institutional Review Board at Boston University Medical Center (Boston, MA).

FHS phenotype and demographic covariate ascertainment

At the 8th in-person examination visit participants completed a questionnaire that inquired about their education, occupation, smoking status, and disease status. Highest levels of educational attainment was assessed by 8 categories—no schooling, grades 1-8, grades 9-11, completed high school or GED, some college but no degree, technical school certificate, associate degree, Bachelor's degree, graduate or professional degree. These categories were collapsed into (1) high school degree or less, (2) some college, and (3) bachelor's or graduate degree for the current analysis. Smoking status was separated into current smokers, former smokers, and never smokers. Diabetes was defined as having fasting blood glucose ≥126 mg/dl or current treatment for diabetes. Cardiovascular disease was determined by a panel of 3 physicians, who reviewed participants' medical records, laboratory findings, and clinic exam notes.

FHS DNA methylation measurement

Peripheral blood samples were collected at the 8th examination (2005-2008). Genomic DNA was extracted from buffy coat using the Gentra Puregene DNA extraction kit (Qiagen, Valencia, CA) and bisulfite converted and cleaned using EZ DNA Methylation kit (Zymo Research Corporation, Irvine, CA) according to manufacturer's instructions. DNA methylation quantification was conducted on the Infinium Human Methylation450 Beadchip (Illumina, San Diego, CA). Paired samples were bisulfite-treated and hybridized together. The pairs were then randomly distributed across Infinium BeadChips and hybridization dates.

Statistical analyses

CAD associations

Demographic characteristics of the total, filtered, and paired study samples were summarized in Table 1. Bivariate associations by CAD diagnosis were calculated and compared using student t-tests and linear model trend tests for continuous and categorical data, respectively (Table 2).

Methylation data pre-processing

Raw IDAT image files were processed in R (version 3.1) and methylation beta values were generated using the Bioconductor minfi package (version 1.12.0) with background correction (Aryee et al., 2014). Sample exclusion criteria included poor SNP matching of control positions (n=22), probe missing rate > 10% (n=8), outliers from multi-dimensional scaling (MDS) (n=34), outliers from methylated/unmethylated channel intensity scores (n=10) and sex mismatch (n=2). Probes were excluded if missing rate > 10% (n=684). Probes overlying annotated SNPs in the probe CG site or single-base extension site were also removed (n=23,486). In total, 471 samples and 443,304 CpG probes remained for analysis.

DNA methylation data was stratified-subset quantile normalized and principal components analysis (PCA) was used to assess potential batch effects, particularly of measured covariates such as sample position on array and hybridization date. Batch effects related to hybridization date were removed using ComBat in the sva package (Leek and Storey, 2007). PCA on the ComBat adjusted matrix was used to assess the persistence of batch effects.

Proportions of six major blood cell types (granulocytes, monocytes, B cells, NK cells, CD4T cells, and CD8T cells) were estimated from methylation data using cell type methylation reference panels (Houseman et al., 2012), as implemented through minfi.

Single site discovery

Individual differentially methylated probes (DMPs) were identified using the limma package (version 3.32.1) (Smyth, 2005). Multivariable linear regression was used to identify individual CpG sites that were associated with CAD on the full, unpaired data set (n=448). The primary model was adjusted for age, sex, granulocyte and CD4T estimated cell counts, and plate ID. To estimate standard error, empirical Bayesian methods were used (Smyth, 2004). Permutation-based P values to protect family wise error rates (FWER) across the genome were calculated based on 1,000 permutations of the CAD phenotypes. Similarly, at each CpG site, per-test P values protecting test-wise error rates (TWER) were calculated. Model fit was assessed using P-P plots and estimated lambda inflation factors. Enrichment of gene ontology biological processes was tested using the 1e-3 threshold.

We performed multiple sensitivity analyses across all probes. First we performed a full CAD risk factor multivariable linear regression model on the full dataset (N=448). This model was adjusted for age, sex, cell type, plate, systolic blood pressure, body mass index, smoking status, alcohol consumption, diabetes status, and education level. Secondary analyses also included conditional linear regression on the paired (n=336) data set. This was implemented via the clogit function (Therneau, 1999) with CAD as the outcome variable, stratified by pairs, and adjusted for percent granulocytes and CD4T cells. Likewise we performed a full risk factor conditional logistic regression model that also considered systolic blood pressure, body mass index, smoking status, alcohol consumption, diabetes status, and education level.

Comparison between probe overlaps in primary and secondary were assessed linear regression between probe p-values and beta-values.

DMR discovery

We tested for differentially methylated regions (DMR) based on our primary model using the bumphunter package (version 1.5.5) as previously described (Jaffe et al., 2012). We used 1,000

bootstraps to protect the relationships between covariates. We set a maximum gap between CpGs of 300 bp and queried regions with a quantile (0.995) based threshold.

Results

Study sample characteristics

The association between DNA methylation and CHD was studied in a paired case-control subset of the Framingham Heart Study. After sample removal for quality control, one hundred and sixty-eight CHD cases were paired with controls by sex and age within 5 years. The median age of the cohort was 77 years (IQR = 13) and consisted of 256 men and 90 women (Table 15). Median BMI was 28.26 (IQR = 6). Education was divided into tertiles (GED/post-high school education/Bachelor's and higher degrees) with 123:100:126 in each category. The cohort consisted of 18 current smokers, 241 former smokers, and 87 never-smokers. Heavy alcohol use (greater than 6 drinks per week) was present in 92 individuals, moderate use (1-6 drinks per week) in 167 individuals, and 86 participants did not drink. Finally, doctors diagnosed 65 individuals with diabetes or high blood sugar.

As this was a paired study, there were no associations between sex or age and CHD status. Out of the other covariates, only education and smoking had significant associations with CHD status (p-values 0.0449 and 5.28e-06, respectively, Table 16). BMI, alcohol use and diabetes or high blood sugar all had association p-values above 0.3.

In this study, we identify differentially methylated positions (DMPs) associated with CHD as categorized by at least one of the following criteria: 1) myocardial infarction diagnosed either with electrocardiogram or enzyme assay, 2) percutaneous transluminal coronary angioplasty procedure or 3) coronary artery bypass graft procedure. DNA methylation was assessed using the Illumina 450k HumanMethylation array. Data pre-processing, normalization and quality control procedures are described in Methods.

Single site DNA methylation discovery

We find significant association between multiple CpGs and CHD status with 21,364 probes found to be differentially methylated ($p<0.05$). The top probes and the nearest genes are shown in Table 17. Approximately half (52%) of probes were hypomethylated. Several of the top differentially methylated CpGs are near genes that have previously been implicated in GWAS studies, such as PITX2, FLT1, PPAP2B, STK32B, and others are near closely related genes, such as PHACTR2 and PHACTR3 for the GWAS PHACTR1 loci.

Regional DNA methylation discovery

As neighboring CpGs are known to have correlated methylation (Huynh et al., 2014), and as we have previously found multiple CpGs in specific regions linked to common disease (Multhaup et al., 2015) we identified regions of the genome where multiple neighboring CpGs (within 300bp) were all associated with CHD status. We found 83 regions with area p-values < 0.05. These DMRs were also near several GWAS-implicated genes (APOB, HLA-DRB1). Out of the top 83 DMRs, only 5 were near genes also implicated by the DMP analysis.

Discussion

This work is ongoing and currently in progress. While we find significant DMPs and DMRs, the p-values associated with these do not remain significant after bonferroni and/or FDR multiple testing correction. There is some debate currently in the scientific community about the appropriateness of bonferroni as applied to methylation data. This is because the bonferroni correction assumes that each test is independent, and there is a large body of literature suggesting that CpG methylation levels are actually highly correlated between CpGs when the CpGs are within ~200 bp of each other. A recent Nature Neuroscience paper, for example, wrote “conventional methods for multiple-test correction,

such as those used in genome-wide association studies, are likely to be overly stringent and inappropriate given the non-independence of DNA methylation across multiple CpG sites and lack of inter-individual variation at many loci; in this study, we therefore report nominal *P* values" (Lunnon et al., 2014).

In practice, however, any reasonable publication will have either results that survive some recognized multiple testing correction OR stringent replication in separate cohorts. Therefore, in order to replicate our results, we are currently pursuing replication in three separate cohorts. Code has been provided to ensure that analytical methods are duplicated across cohorts and to reduce variation. Individual probe results from each cohort will be meta-analyzed together and the resulting meta-analysis p-values will be adjusted by bonferroni correction. It is the hope of this author that these results will be available in the near future.

Figures

Figure 1

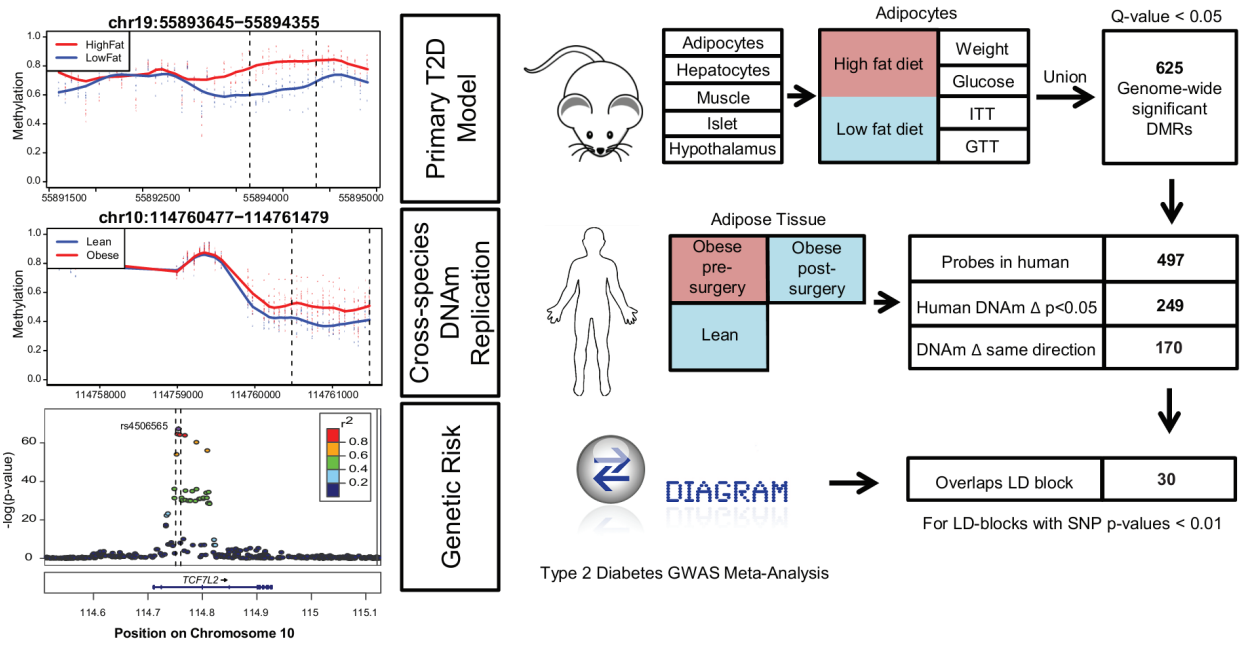

Figure 1**: Diagrammatic explanation of our experimental design and results. Left top and middle panels:** Plots of mouse and human methylation, respectively. Each point represents the methylation level from an individual mouse or human at a specific genomic location, with smoothed lines representing group methylation averages. These points are colored blue for low-fat-fed mice and lean humans and red for high-fat-fed mice and obese humans. The methylation values range from 0 to 1, with 0 equal to minimum detectable methylation and 1 equal to maximum detectable. **Left bottom panel:** plot of significance of SNP association with T2D. Y-axis represents log p-values of the linkage. Bottom axis shows genomic location.

Figure 2

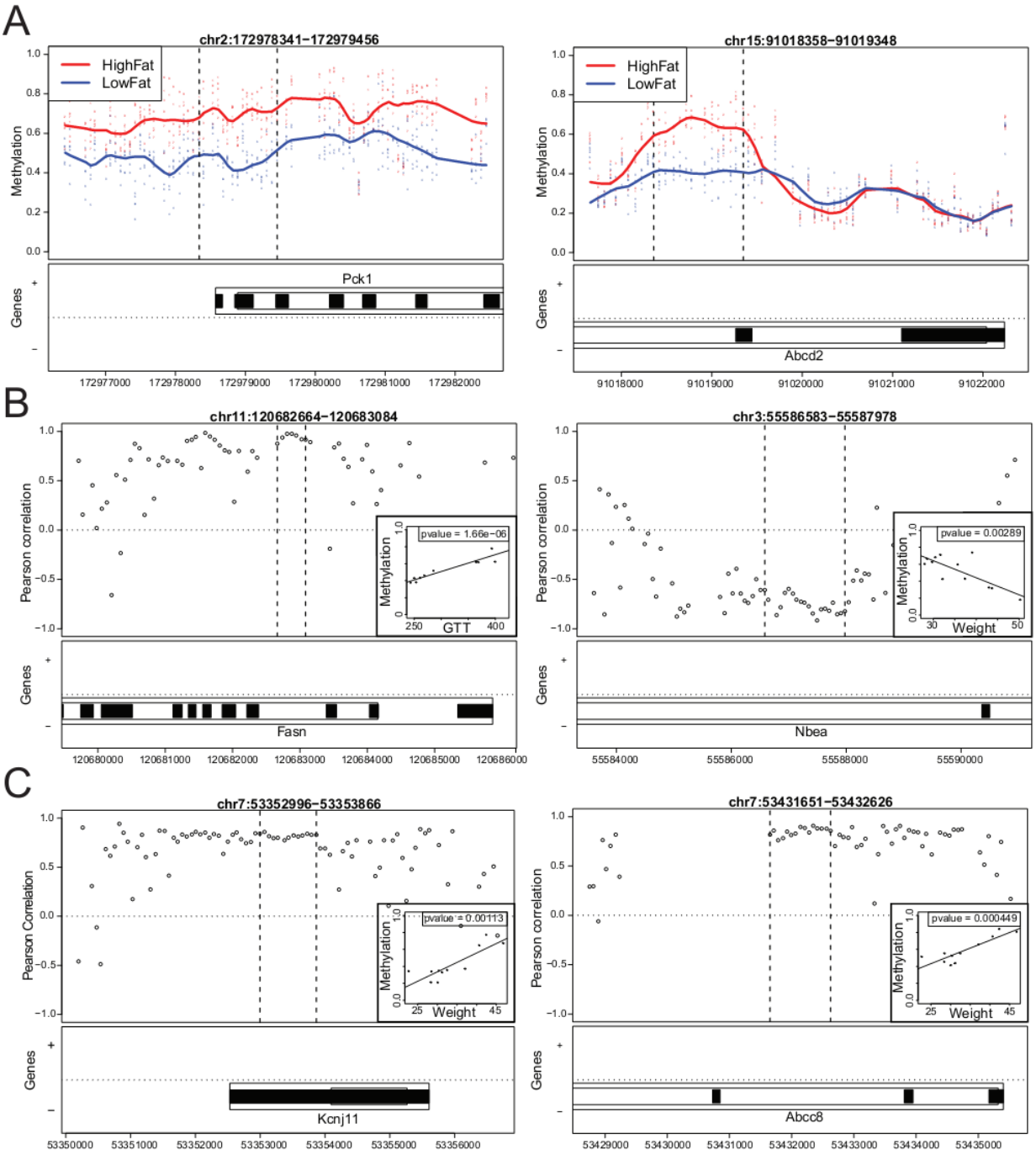

Figure 2**. Genome-wide significant methylation changes related to diet-induced obesity in C57BL/6 mice. (A)** Two genome-wide significant DMRs located at *Pck1* (left) and *Abcd2* (right) are hypermethylated in adipocytes purified from mice raised on a high-fat diet. Each point represents the methylation level in adipocytes from an individual mouse at a specific genomic location, with smoothed

lines representing group methylation averages. These points are colored blue for lean mice and red for obese mice. **(B)** Body weight (grams) and glucose tolerance (AUC) are associated with methylation in adipocytes at genome-wide significant levels at the *Fasn* (left) and *Nbea* (right) genes, respectively. Each point in the top panels represents one probe, with the y-axis representing the Pearson correlation coefficients of the probes with the analyzed phenotype. Dotted lines represent the extent of the DMR as generated automatically via CHARM. The bottom panels display gene location information for the chromosomal coordinates on the x-axis. **(C)** Body weight (grams) is positively associated with methylation in pancreatic islets at genome-wide significant levels at the *Kcnj11* (left) and *Abcc8* (right) genes. For all insets, the y-axis represents methylation, the x-axis displays the specific phenotype associated with methylation in the larger plot, and each point represents the association of the average methylation for all probes in the DMR in one sample with the corresponding sample phenotype.

Figure 3

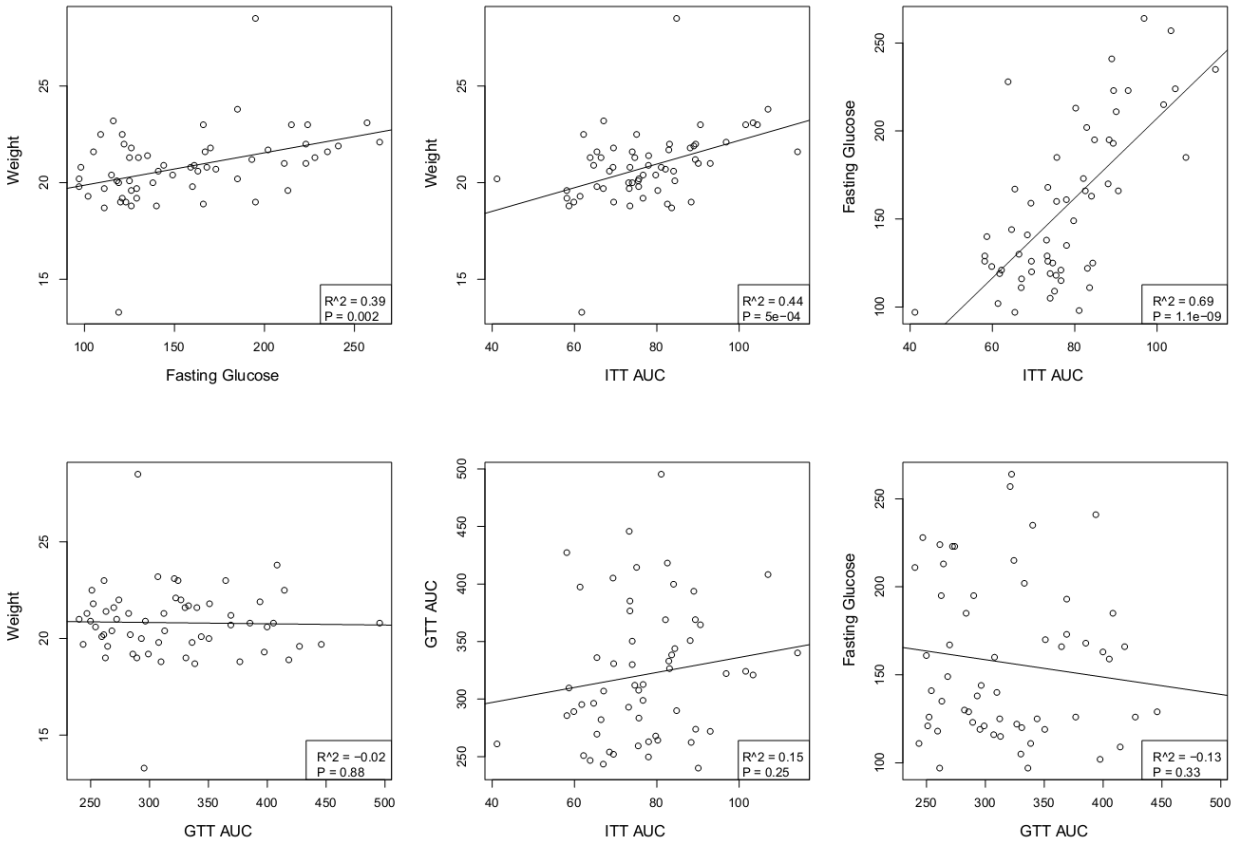

Figure 3**. Correlation of metabolic traits in diet-induced obesity mouse model.** Correlations between the mouse traits observed over time. Mouse weight, fasting glucose levels (collected at the time of glucose tolerance test), and insulin tolerance test and glucose tolerance test area-under-the-curve scores are plotted and correlated against each other. Correlation coefficients and p-values for the linear models are shown in the inserts.

Figure 4

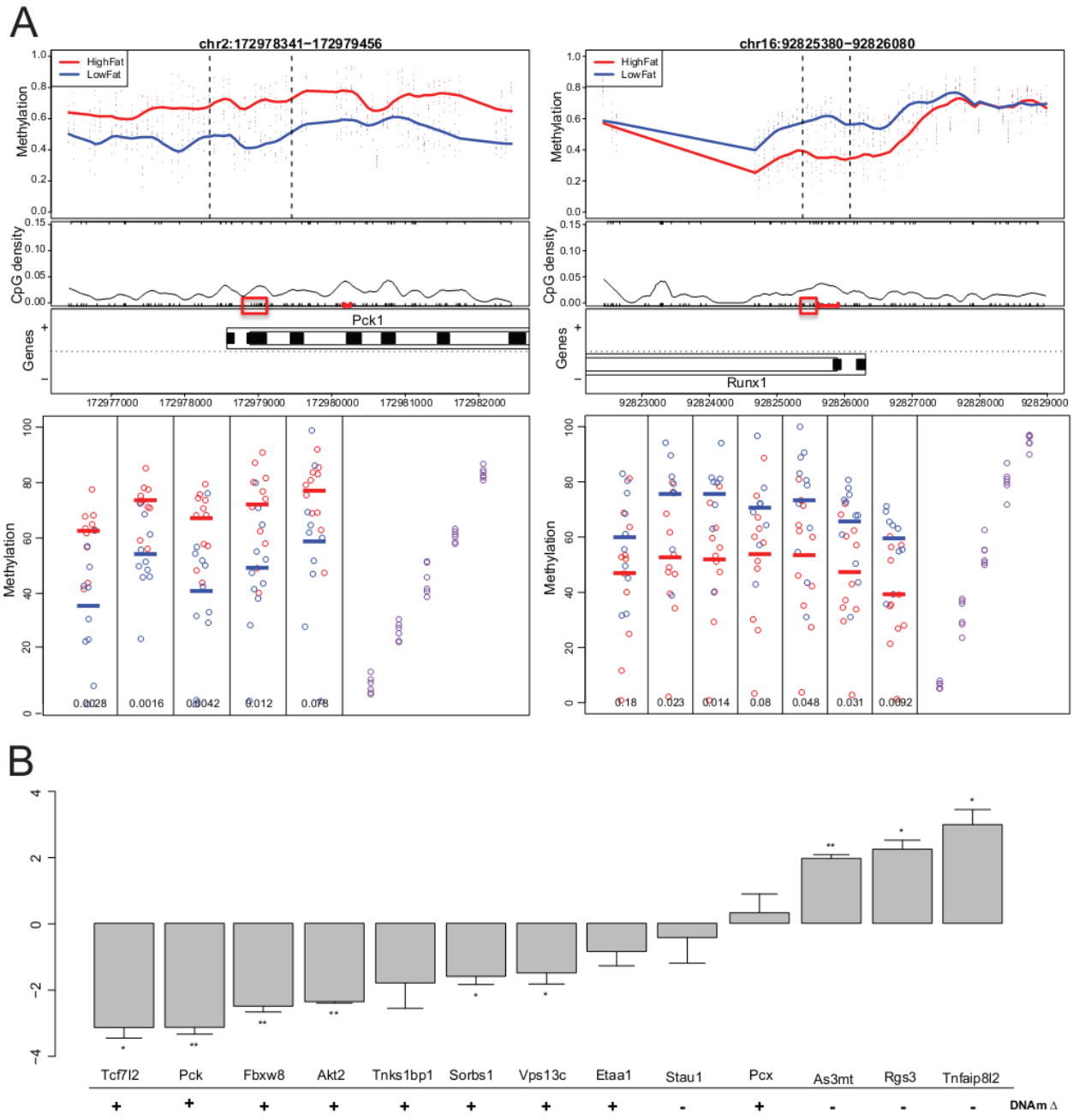

Figure 4**. Replication of mouse methylation changes in additional mice, and associated gene expression changes. (A)** Methylation changes observed after CHARM analysis at two genome-wide significant DMRs located at *Pck1* (left) and *Runx1* (right) are replicated in independent samples from 9 high-fat and 9 low-fat mice using bisulfite pyrosequencing. Red boxes indicate CpGs assayed in pyrosequencing. For the lower pyrosequencing plots, the y-axis represents methylation, and individual

CpGs are plotted along the x-axis. Purple dots represent controls artificially methylated to have 0, 25, 50, 75 and 100% methylation. **(B)** Gene expression changes for genes near genome-wide significant mouse adipocyte DMRs. RNA levels were normalized to same-sample 18S RNA measurements and are displayed as $[C_T$ (high-fat samples) – C_T (low-fat samples)$]^2$. The direction of the genome-wide significant CHARM DMR closest to the gene is denoted below the gene names; + and − represent regions hyper- or hypomethylated in the high-fat samples, respectively.

Figure 5

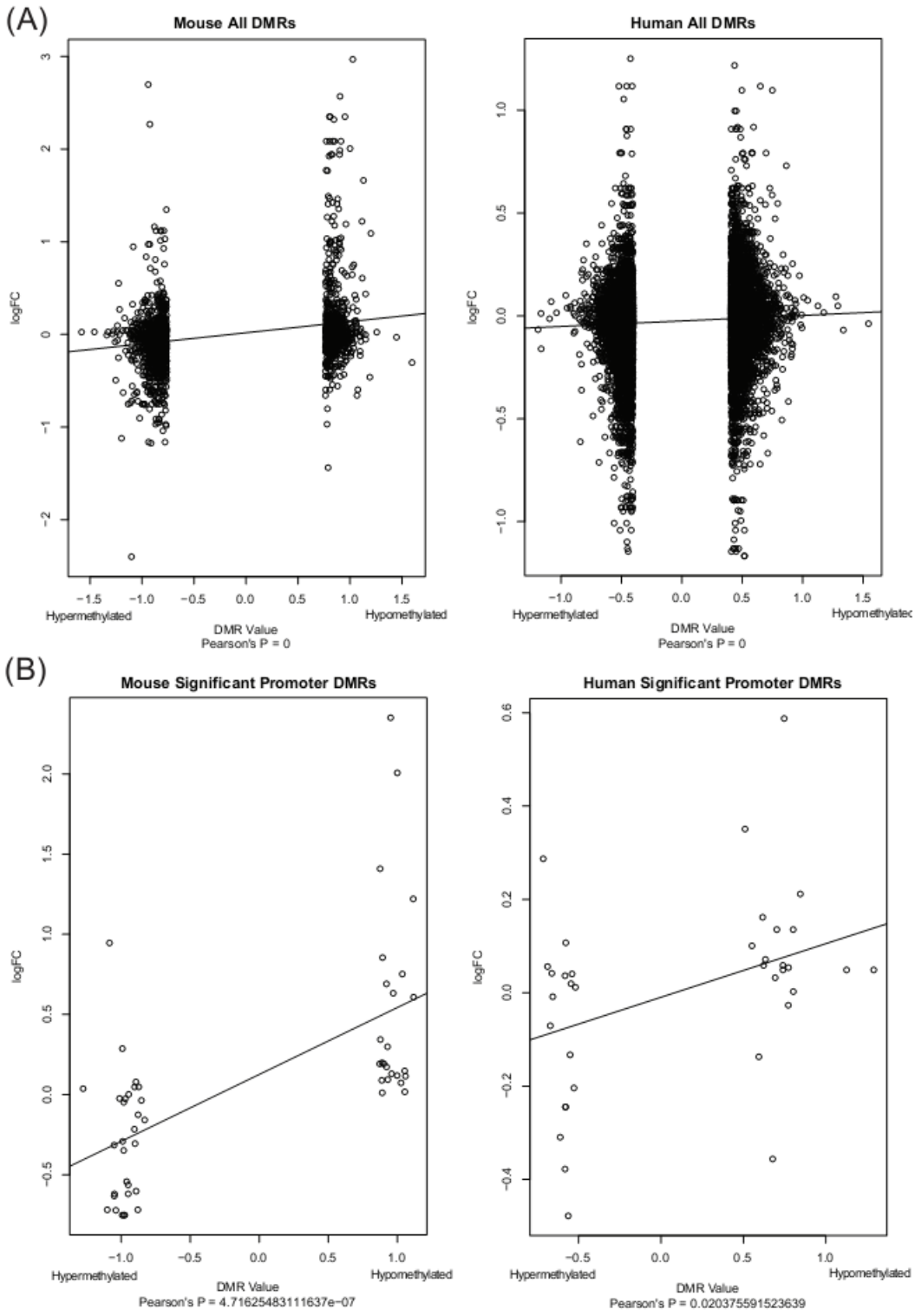

Figure 5**. Correlation of methylation and gene expression in mouse and human adipose tissue.** This figure shows the relationship between methylation and gene expression in both mouse and human

adipose tissues. Gene expression data was downloaded from GEO (see Materials and Methods) and plotted against mouse adipocyte and human adipose tissue CHARM data. Y-axes are the logarithm of the fold change (logFC) of the gene expression in high-fat-fed mice and obese humans versus low-fat-fed mice and lean humans. X-axes are the DNA methylation values calculated by CHARM (Table S1) for the high-fat versus low-fat mouse and obese versus lean human comparisons. Here, higher values indicate hypomethylation in high-fat / obese samples. P-values are for Pearson product-moment correlations versus a null hypothesis of no correlation.

Figure 6

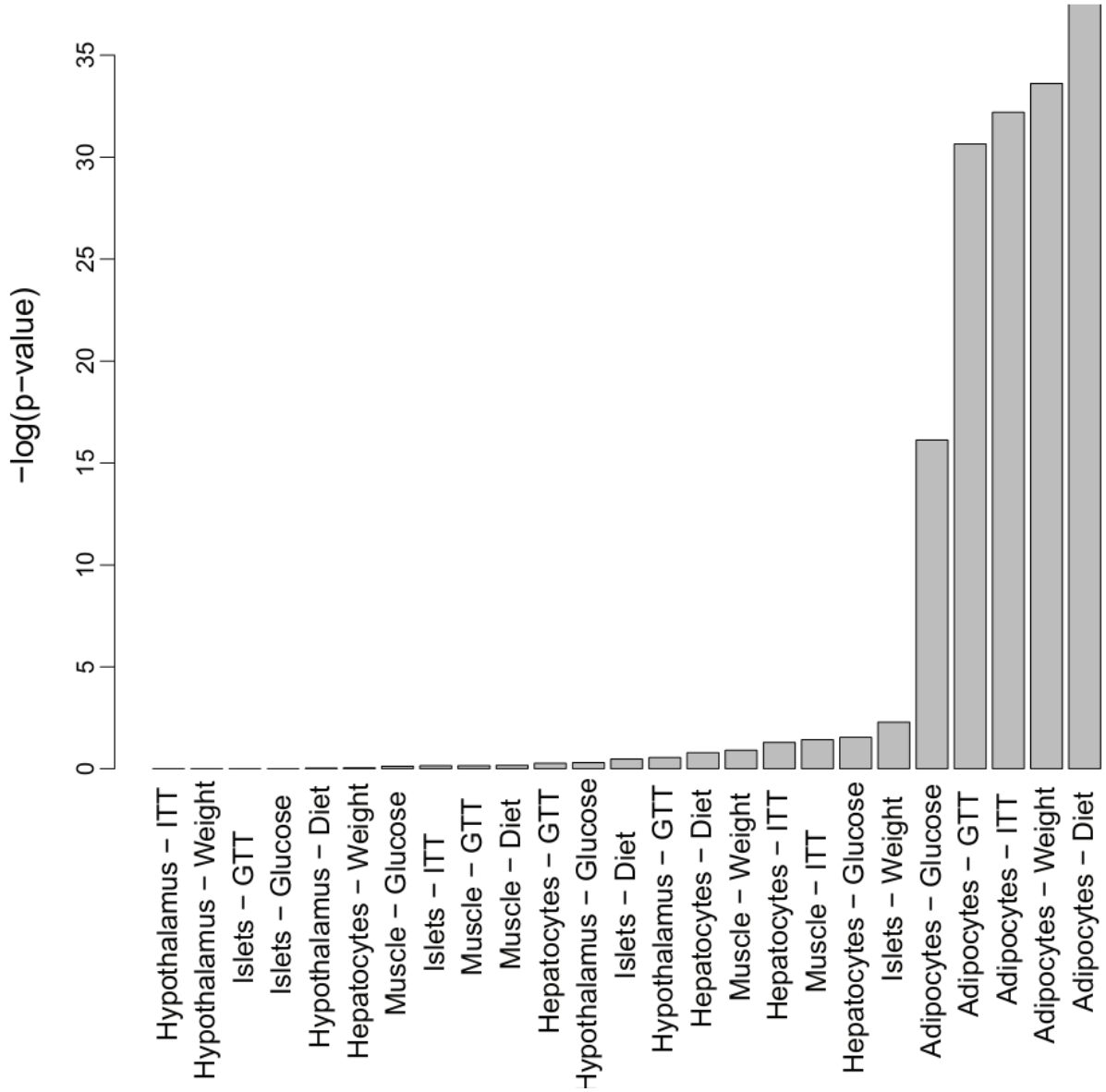

Figure 6**. Significance of methylation change overlap between mouse and human tissues.** In this figure, all 25 mouse analyses (x-axis) are compared against the human adipose obesity analysis. Values plotted represent the largest –log(p-value) for chi-squared tests for the overlap for all DMRs with nominal p-values < 0.05 between the given mouse analysis and the human adipose obesity analysis.

Figure 7

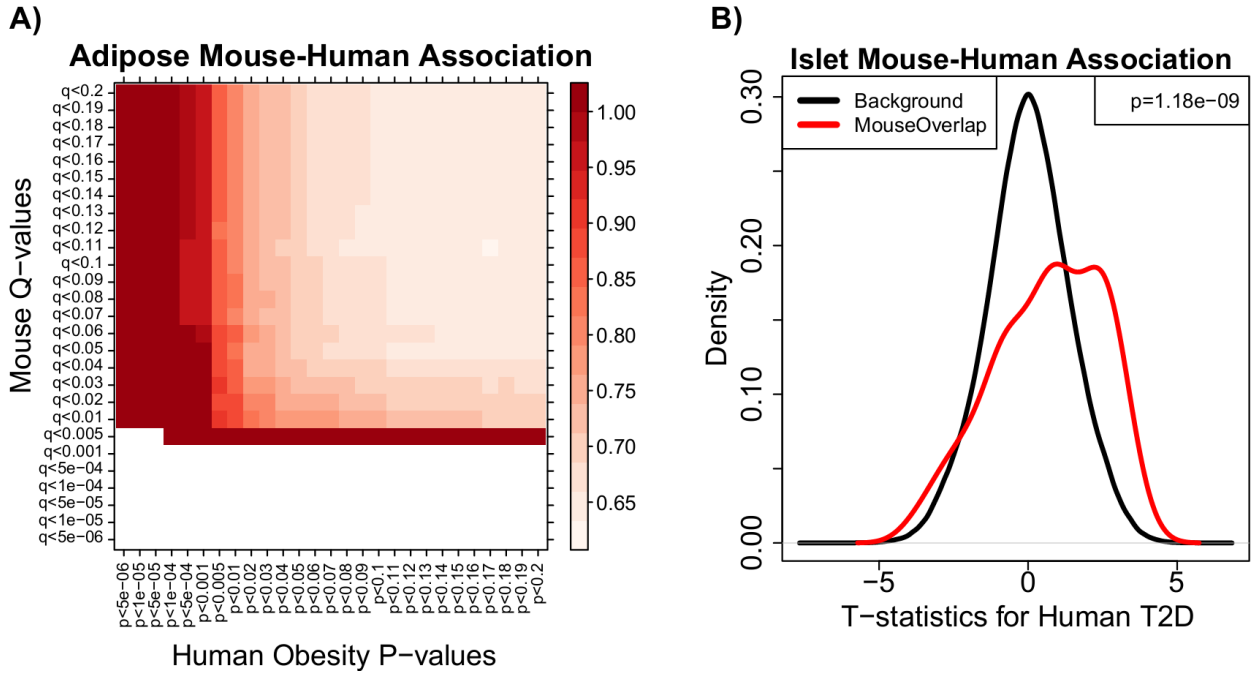

Figure 7**. Consistent mouse-human methylation changes. A)** For each square, the proportion of conserved mouse and human regions that had directionally consistent methylation changes in adipose tissue between species was calculated. Regions were required to have mouse and human methylation changes at or below the indicated Q-value for mouse and P-value for human. The color indicates the proportion of directionally consistent regions, with darker colors indicating a higher proportion. B) The observed versus expected T-statistics for the proportion of overlap between the CHARM pancreatic islet mouse methylation data and the previously reported Illumina Infinium 450k BeadChip pancreatic islet human methylation data.

Figure 8

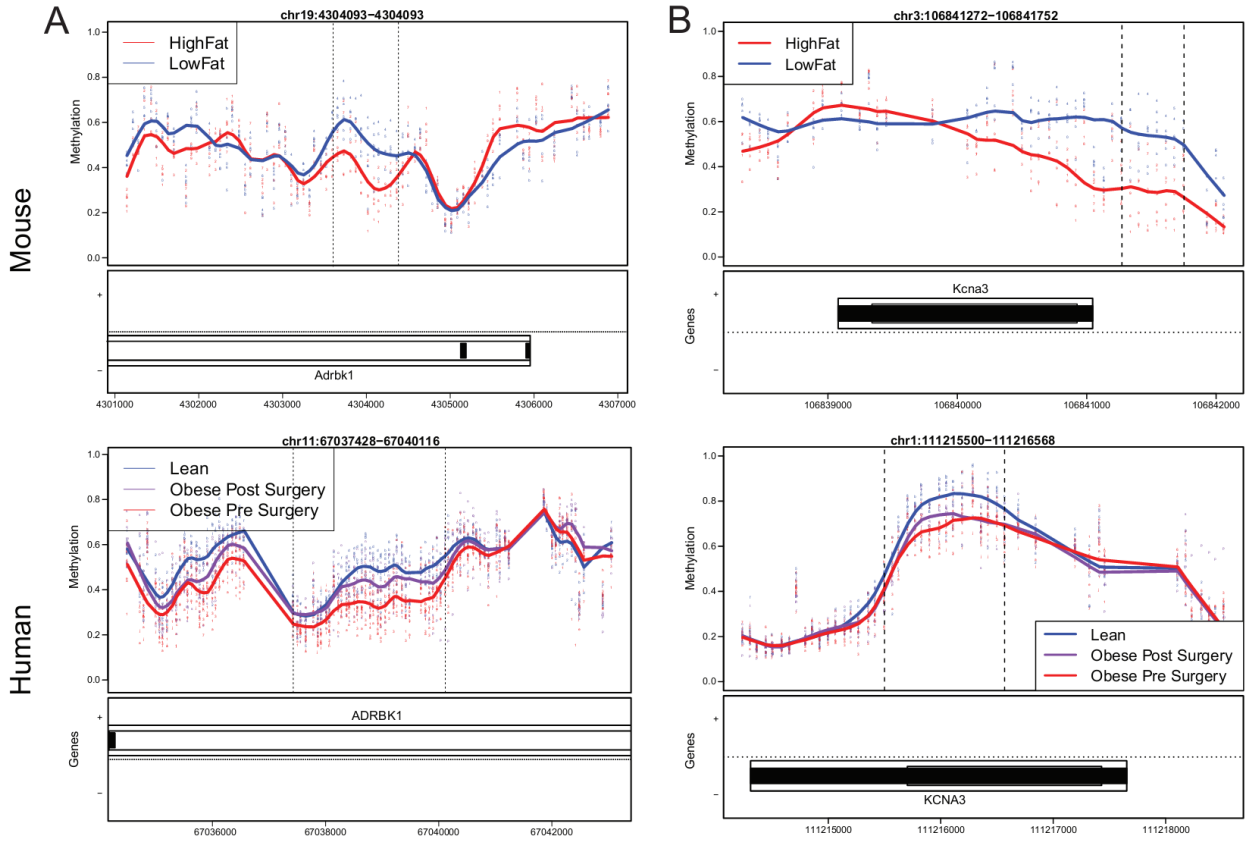

Figure 8**. Overlapping methylation changes in human and mouse adipose tissue.** Two genome-wide significant DMRs found in mouse adipocytes over *Adrbk1* **(A**, top**)** and *Kcna3* **(B**, top**)** are shown along with the corresponding methylation changes in human adipose tissue in **(A**, bottom**)**, and **(B**, bottom**)**. For the panels denoting methylation, each point represents the methylation level from an individual mouse or human at a specific genomic location, with smoothed lines representing group methylation averages. These points are colored blue for low-fat-fed mice and lean humans and red for high-fat fed mice and obese pre-surgery humans. Purple denotes samples from obese humans post-RYGB surgery. The methylation values range from 0 to 1. Below each methylation plot is a panel showing genomic coordinates for the respective species and any genes at those coordinates.

Figure 9

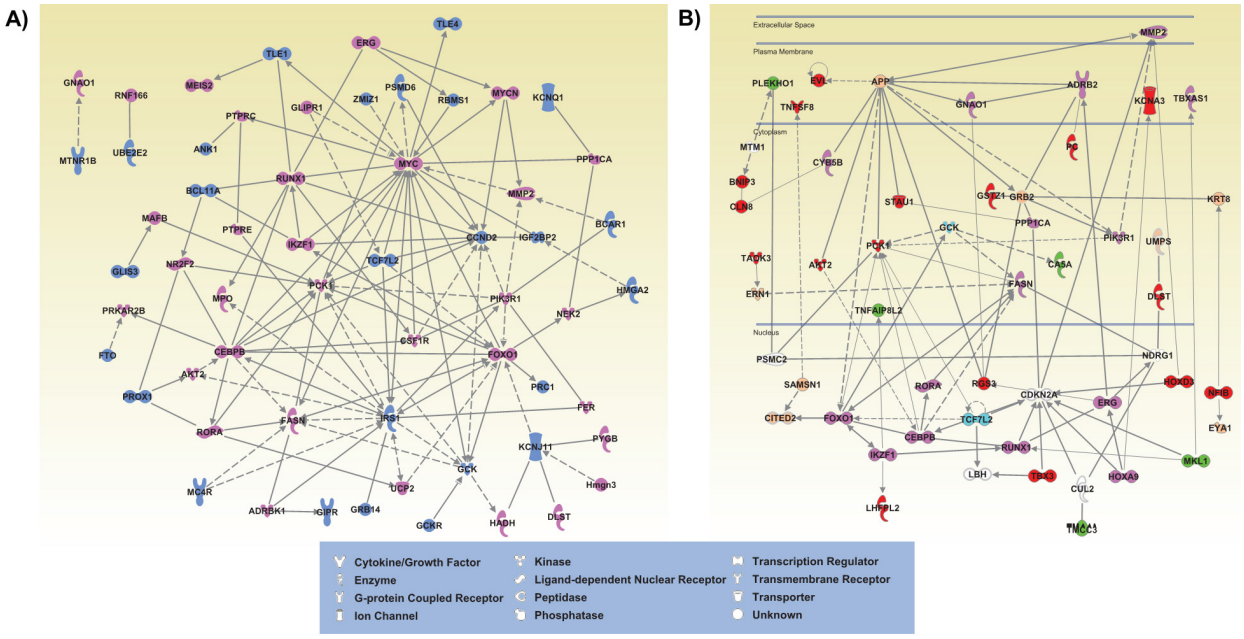

Figure 9**. Diagrammatic representation of the interactions between epigenetically conserved and genetically-associated genes implicated in this study.** Generated using QIAGEN's Ingenuity IPA (Ingenuity® Systems, www.ingenuity.com), these diagrams represent the connections between genes implicated in our analysis of adipocytes and adipose tissue. A) All genes with genome-wide significant linkage to T2D in the DIAGRAM meta-analysis were connected to genes near directionally and locationally conserved DMRs. Genes with no connections were dropped. B) Starting with the set of 30 genes near directionally conserved cross-species DMRs that also overlapped genes with nominally significant T2D GWAS association, this network was grown by adding genes near species-conserved and mouse-only genome-wide significant DMRs in order to represent one potential regulatory network. The genes are colored as follows: **Red**—genes showing conserved differential methylation in mouse and human diet-induced diabetic phenotypes, with genetic T2D risk loci association (Table 3); **Green**—genes with a possible functional role in insulin sensitivity assayed by insulin-stimulated glucose uptake in this paper (Figure 6); **Blue**—genes from the DIAGRAM meta-analysis that are genome-wide significant;

Purple—genome-wide significant mouse DMRs that were conserved and directionally consistent in human obesity (Table S4); **Orange**—genome-wide significant mouse DMRs that are not also conserved in human (Table S1); **White**—genes added for annotation clarity, not present in our results.

Figure 10

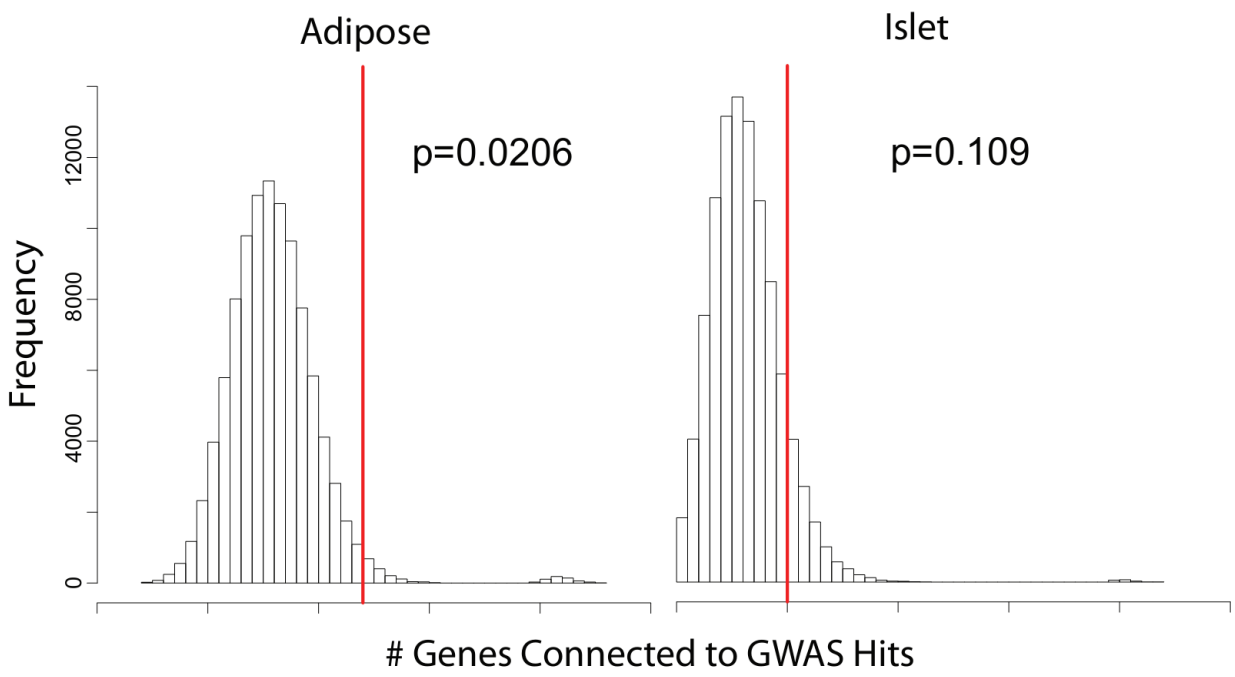

Figure 10**. Enrichment of connections between genes implicated by methylation and genome-wide significant GWAS genes.** This figure shows expected and observed connections and (both direct protein interactions and transcriptional control) and overlap between genes near species-conserved adipose and islet DMRs and genes with genome-wide significant linkage to T2D in the DIAGRAM GWAS meta-analysis**.** The set of all possible one-step connections to the DIAGRAM GWAS genes was pulled from the Ingenuity Knowledge Base, and the GWAS genes themselves were added. 100,000 permutations of random genes near DMRs were overlapped with this set, and the number of overlaps from the permutations are represented by the histograms. The actual number of observed DMR-GWAS connections is denoted by the vertical red line, and the p-values represent permutation p-values for the difference between observed and expected connections.

Figure 11

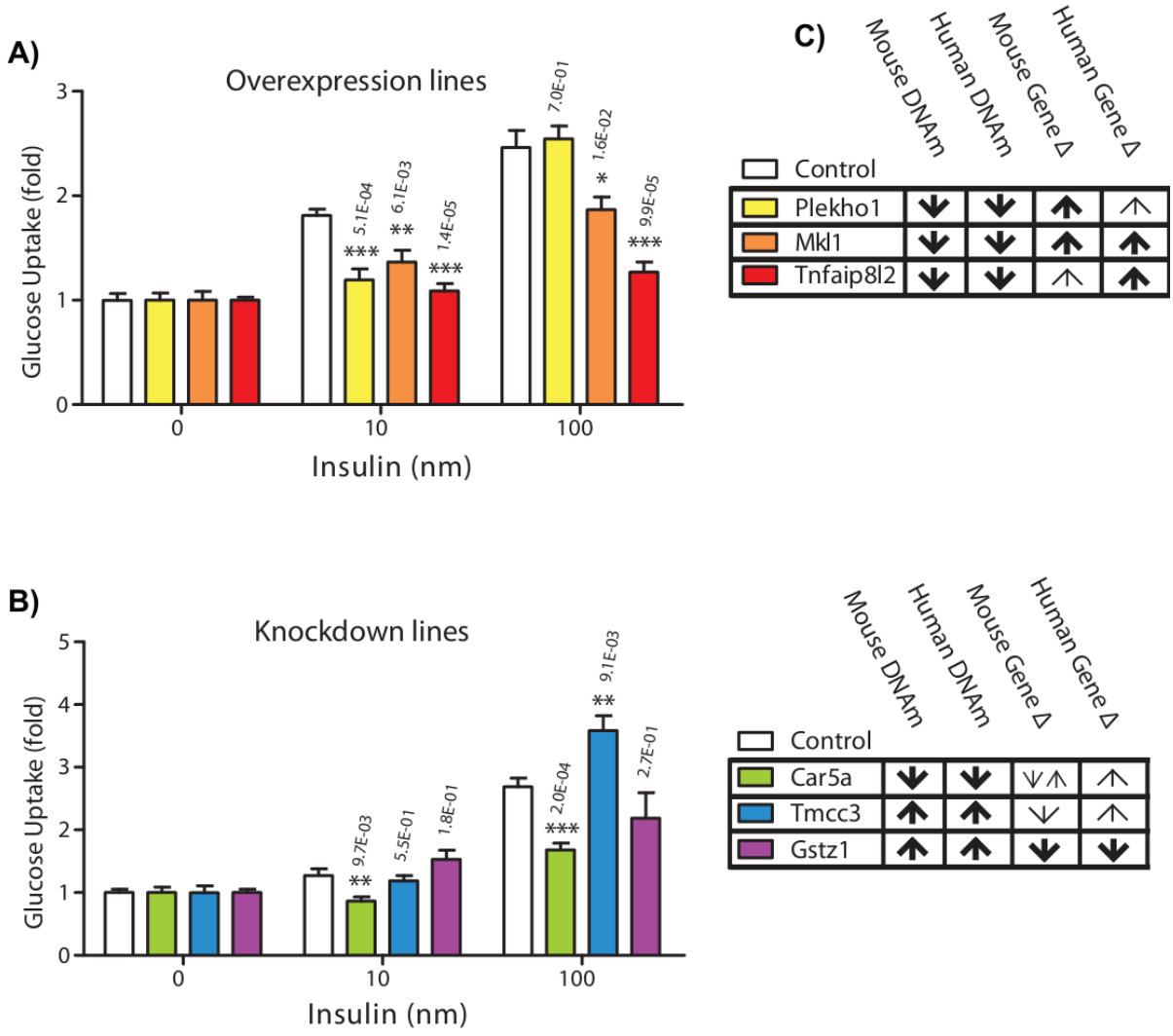

Figure 11**. Overexpression and shRNA-mediated knock down of selected genes in 3T3-L1 adipocytes.** Selected genes from the set of 30 species conserved and T2D-SNP overlapping adipose DMRs were either stably overexpressed **(A)** or knocked down with shRNA **(B)** as noted. After differentiation into mature adipocytes, these cell lines were then subjected to insulin-stimulated glucose uptake assay. Glucose uptake is plotted as fold difference from normal, and significance was determined by two-way ANOVA modified by Bonferroni correction denoted as follows: * p<0.05, ** p<0.01, *** p<0.001**. (C)** DNA methylation and gene expression levels for high-fat-fed mice and obese human versus low-fat-fed

mice and lean humans (e.g., "↓" indicates hypomethylation / lower gene expression in high-fat-fed and obese compared to low-fat-fed and lean). **Bold arrows** indicate significant changes; non-bold arrows are non-significant changes.

Tables

Table 1

Table 1**. Genome-wide significant mouse DMRs**

Tissue	Analysis	Q-val < 0.05	Q-val < 0.1
Adipocytes	Diet	232	448
	Weight	183	288
	Fasting Glucose	235	571
	GTT	0	3
	ITT	294	419
Islets	Diet	4	4
	Weight	446	495
	Fasting Glucose	0	0
	GTT	0	0
	ITT	0	0
Hepatocytes	Diet	0	0
	Weight	6	25
	Fasting Glucose	0	0
	GTT	0	0
	ITT	2	2
Hypothalamus	All Comparisons	0	0
Muscle	All Comparisons	0	0

Q-values generated based upon comparison of observed DMR areas to areas generated by 1000 random permutations of phenotype/methylation associations.

Table 2. See Appendix 1

Table 2**. Mouse comprehensive high-throughput array-based relative methylation (CHARM) results, related to Table 1.** These tables show the results of our CHARM analysis for the five assayed mouse tissues against the five measured metabolic phenotypes of diet, fasting glucose, mouse weight, glucose tolerance test and insulin tolerance test. Listed in each table are the genomic coordinates for each differentially methylated region (DMR), the average value of the smoothed effect estimate within the DMR (value), the number of probes in the DMR (nprobes), the statistical area (nprobes * value, area), a p-value for the difference in methylation at each DMR (pval_pool), family-wise error rate-corrected p-values (pval_fwer), q-values derived from false discovery rate (see methods, qval_pool), the gene symbol and RefSeq annotation of the nearest gene, and multiple parameters describing the physical relationship of the DMR to the gene in the genome. As "% change" we also note the largest change in beta-value for any CHARM probe within the listed DMR. These are roughly the inverse logit of the scores from the value column with the surrogate variables regressed out, and are the numbers used to generate our plots. For diet analyses, these were differences between the average of samples from the low-fat and high-fat groups, and for continuous variable analysis these were differences between the average of the tertile of samples with the highest variable scores and the tertile with the lowest scores.

Table 3

Chr	Start	End	Width	qval	Gene	relation	distance
chr1	25111126	25111547	422	0.015	Clic4	inside	26857
chr1	64061530	64062519	990	0.006	Pgm2	inside	1768
chr1	1.01E+08	1.01E+08	465	0.005	Dbt	inside	1015
chr1	2.3E+08	2.3E+08	415	0.026	Galnt2	inside	33499
chr2	1.69E+08	1.69E+08	160	0.042	Lass6	inside	117391
chr8	17529635	17529755	121	0.047	Mtus1	inside	28971
chr10	97517128	97517433	306	0.032	Entpd1	inside	1149
chr10	1.3E+08	1.3E+08	447	0.005	Ptpre	inside	114341
chr12	11812172	11812279	108	0.022	Etv6	inside	9226
chr12	1.25E+08	1.25E+08	470	0.006	Ncor2	inside	126377
chr12	1.26E+08	1.26E+08	234	0.035	Aacs	inside	19023
chr17	40142289	40142472	184	0.032	Dnajc7	inside	27020
chr17	48838267	48838764	498	0.036	3300001P08Rik	downstream	35690
chr5	34655547	34655935	389	0.029	Rai14	promoter	1628
chr5	34687673	34687965	293	0.029	Rai14	inside	23832
chr5	77803248	77803367	120	0.03	Lhfpl2	inside	118687
chr5	1.07E+08	1.07E+08	257	0.011	Efna5	inside	96616
chr5	1.07E+08	1.07E+08	438	0.006	Efna5	inside	64136
chr7	47672033	47673452	1420	0.006	Tns3	upstream	24418
chr7	1.51E+08	1.51E+08	177	0.035	Nub1	inside	553
chr14	89707672	89708616	945	0.009	Foxn3	inside	192777
chr3	72870608	72870903	296	0.036	Shq1	inside	25788
chr3	72896699	72897027	329	0.047	Shq1	inside	466
chr16	30455438	30455975	538	0.005	Sephs2	inside	1221
chr16	56137894	56138256	363	0.006	Gnao1	upstream	85203
chr16	85376056	85376680	625	0.005	Gse1	upstream	144355
chr21	37412563	37413036	474	0.006	Setd4	inside	17406
chr11	27473339	27473770	432	0.009	Lgr4	inside	18196
chr11	27490752	27491523	772	0.01	Lgr4	inside	2692
chr11	57067776	57068975	1200	0.005	Tnks1bp1	inside	20866
chr15	61136559	61136807	249	0.024	Rora	inside	381025
chr15	62343394	62344034	641	0.005	Vps13c	inside	7451
chr4	1.14E+08	1.14E+08	773	0.005	Larp7	downstream	232193

Table 3**. Adipocyte DMRs with genes in common between adipocyte and pancreatic islet analyses.**

These are the adipocyte DMRs with Q-values < 0.05 from Table 1 that are near genes that were also implicated by pancreatic islet DMRs with Q-values < 0.05.

Table 4

Table 4**: Gene ontology for genes near DMRs**

DMR set	Gene Ontology Biological Process Term	Count	List Total	Enrichment Q-value
Adipocyte Hypermethylated DMRs	triglyceride metabolic process	9	47	3.33E-04
	acylglycerol metabolic process	9	56	8.23E-04
	neutral lipid metabolic process	9	58	7.51E-04
	positive regulation of cholesterol storage	4	6	2.15E-03
	low-density lipoprotein particle clearance	4	7	3.96E-03
	cellular carbohydrate metabolic process	11	131	9.12E-03
Adipocyte Hypomethylated DMRs	regulation of response to stimulus	81	2173	3.13E-05
	neutrophil chemotaxis	8	32	7.50E-04
	positive regulation of response to stimulus	48	1110	5.10E-04
	cell activation	23	334	4.70E-04
	myeloid leukocyte activation	10	59	4.13E-04
	immune system process	38	784	3.70E-04
Pancreatic Islet DMRs	regulation of cellular process	160	7177	4.70E-02
	regulation of biological process	166	7610	5.20E-02
	regulation of peptide hormone secretion	11	129	5.15E-02
	regulation of peptide secretion	11	134	5.54E-02
	regulation of insulin secretion	10	111	4.74E-02
	single-organism cellular process	179	8479	4.16E-02

Genes near genome-wide significant DMRs (q-value <0.05) for adipocyte/fasting glucose and pancreatic islet/body weight associations were submitted to the **G**ene **O**ntology en**RI**chment ana**L**ysis and visua**L**iz**A**tion tool (GOrilla) along with a background of all the genes possible to find on the applicable array. The list of genes found in adipocytes was first divided into hypomethylated and hypermethylated groups depending on the status of the corresponding DMR. Here, hypermethylation refers to areas where increased methylation is associated with higher fasting glucose and hypomethylation the converse.

Table 5

Tissue	Nearest Gene	CHARM High-Fat	CHARM Low-Fat	Pyro CpG 1 High-Fat	Pyro CpG 1 Low Fat
Adipocytes	Chpt1	78.88	44.32	93.3	66.04
Adipocytes	Stap1	32.16	55.37	36.16	51.2
Adipocytes	Pscdbp	38.44	68.24	46.8	66.3
Adipocytes	Pck1	73.23	45.6	60.9	33.7
Adipocytes	Runx1	31.47	58.93	45.13	58.31
Adipocytes	Fasn	77.57	46.63	66.09	51.24
Adipocytes	Dguok	69.12	47.16	63.8	54.87
Adipocytes	Liph	71.81	55.24	66.5	43.88
Adipocytes	Axin2	58.6	36.98	56.43	34.28

Pyro CpG 1 P-value	Pyro CpG 2 High-Fat	Pyro CpG 2 Low Fat	Pyro CpG 2 P-value	Pyro CpG 3 High-Fat
0.029	88.6	54.5	0.017	90.72
0.03	40.67	67.12	0.0042	48.81
0.03	52.4	78.8	0.019	54.09
0.0019	71.95	52.33	0.0012	65.2
0.18	50.8	73.94	0.026	50.21

0.32	**70.95**	**53.14**	**0.24**	**83.24**
0.26	**55.5**	**36.4**	**0.03**	**49.8**
0.039	**77.56**	**44.87**	**0.0077**	**59.57**
0.027	**50.5**	**33.58**	**0.17**	**50.1**

Pyro CpG 3 Low Fat	**Pyro CpG 3 P-value**	**Pyro CpG 4 High-Fat**	**Pyro CpG 4 Low Fat**	**Pyro CpG 4 P-value**
54.47	**0.0089**	**69.7**	**42.03**	**0.039**
71.46	**0.014**	**50.98**	**76.35**	**0.015**
80.1	**0.012**	**55.68**	**78.242**	**0.025**
38.9	**0.003**	**70.1**	**47.6**	**0.011**
73.91	**0.018**	**52**	**69.03**	**0.099**
47.03	**0.022**	**95.05**	**81.26**	**0.22**
37.55	**0.093**	**63.8**	**41.2**	**0.0069**
34.74	**0.026**	**67.88**	**41.17**	**0.015**
25.4	**0.064**	**37.95**	**30.88**	**0.52**

Pyro CpG 5 High-Fat	**Pyro CpG 5 Low Fat**	**Pyro CpG 5 P-value**	**Pyro CpG 6 High-Fat**	**Pyro CpG 6 Low Fat**

65.52	**34.58**	**0.038**	**73.86**	**35.44**
43.96	**66.49**	**0.012**		
75.27	**56.88**	**0.061**		
51.6	**71.5**	**0.048**	**45.56**	**63.8**
75.27	**56.88**	**0.061**	**55.53**	**36.41**
59.75	**36.68**	**0.062**	**55.33**	**35.89**
58.09	**41.36**	**0.069**	**68.91**	**80.2**

Pyro CpG 6 P-value	**Pyro CpG 7 High-Fat**	**Pyro CpG 7 Low Fat**	**Pyro CpG 7 P-value**	**Pyro CpG 8 High-Fat**
0.018	**73.94**	**45.8**	**0.077**	**68.63**
0.03	**37.54**	**57.65**	**0.011**	
0.03	**49.88**	**37.55**	**0.094**	
0.15	**64.49**	**40.96**	**0.076**	**59.85**
0.18				

Pyro CpG 8 Low	**Pyro CpG 8 P-**

Fat	value
46.06	0.13
37.842	0.067

Tissue	Nearest Gene (Mouse)	CHARM value (derived from regression beta)	CpG 1 Pyrosequencing Slope
Hepatocyte	Scd1	0.095453437	0.4194
Hepatocyte	Fermt2	-0.07619625	-0.0294
Hepatocyte	Atp6v0a1	-0.074216772	-0.02607
Hepatocyte	Arhgap29	-0.073105365	-0.2179
Hepatocyte	Scd1	0.076948689	0.1306
Hepatocyte	Masp1	-0.078898672	-0.0006339
Hepatocyte	Elovl5	0.067301744	0.2199
Hepatocyte	Tmem140	-0.083317234	0.01515

CpG 1 Pyrosequencing	CpG 2 Pyrosequencing	CpG 2 Pyrosequencing	CpG 3 Pyrosequencing

p-value	Slope	p-value	Slope
0.016	**0.6214**	**0.026**	
0.38			
0.024	**0.1307**	**0.43**	
0.012	**-0.0122**	**0.82**	
0.018	**0.003109**	**0.95**	**0.244**
0.68	**-0.0008258**	**0.9**	
0.089	**0.2955**	**0.025**	**0.282**
0.87			

CpG 3 Pyrosequencing p-value
0.036
0.042

Table 5**. Mouse replication results.** This table contains the results of our pyrosequencing assays to replicate our CHARM results in separate samples. For adipocyte measurements, this table presents the

levels of methylation in mouse CHARM, corresponding pyrosequencing results, and the p-value corresponding to a two-tailed t-test for difference between the pyrosequencing methylation values from high-fat-fed and normal chow-fed mice. For the hepatocyte results, this table presents the average value of the smoothed effect estimate for weight versus methylation within the DMR in mouse CHARM, the slope of the correlation between methylation and the outcome derived from pyrosequencing in conserved regions from corresponding human samples, and the pearson's p-value for that slope.

Table 6

Table 6. **Tissue Purification qPCR results**			
Gene	**Fold change (high fat/low fat)**	**SD (SD_high_fat - SD_low_fat)**	**Pval (true fold change is not 0)**
Adipoq	**-1.41**	**0.4**	**0.039**
Pparg	**-0.58**	**0.3**	**0.167**
Ccl2	**1.29**	**0.32**	**0.043**
Albumin	**-0.08**	**1.32**	**0.476**
CD68	**1.69**	**0.9**	**0.174**
Coro1a	**0.02**	**1.41**	**0.494**
F4/80	**0.06**	**0.64**	**0.463**
CD14	**0.27**	**0.38**	**0.341**

Table 6**. Results of qPCR assay to test adipose tissue purification, related to Figure 3**. This table shows the results of our quantitative PCR assay to test if our mouse adipocyte tissue samples were pure.

Table 7

Table 7**. Conserved mouse/human adipose tissue DMRs**

Chr	Start	End	Human p-val	Human beta
chr1	6804060	6804159	0.05396806	-0.127370809
chr1	19804495	19805076	0.009584786	-0.102400227
chr1	23171771	23172254	0.163730493	-0.037281993
chr1	27681877	27682185	0.017716616	-0.077604526
chr1	36191885	36192039	0.239591101	-0.036206241
chr1	37838402	37839079	0.378536274	-0.015552746
chr1	48351513	48351750	0.000271516	0.138474484
chr1	61909979	61910398	0.283317249	0.065328087
chr1	63253431	63254413	0.090799505	0.029828429
chr1	64061530	64062519	0.023274675	0.09603559
chr1	65733701	65734201	0.297952055	-0.036724144
chr1	94153416	94153898	0.641591964	0.012994303
chr1	100112308	100112822	0.191059976	0.020861026
chr1	100436169	100436245	0.045055579	0.053255223
chr1	100713823	100714287	0.026114662	-0.043754768
chr1	111214849	111215337	0.000830083	-0.174664378
chr1	111218226	111218674	0.028609698	0.071486755
chr1	114413127	114414067	8.28E-05	-0.140138463
chr1	114414346	114414954	8.28E-05	-0.140138463
chr1	114415254	114415607	8.28E-05	-0.140138463
chr1	149983089	149983332	0.481806589	0.013695731
chr1	150129607	150129911	0.011515715	-0.061527856
chr1	151129480	151129904	2.58E-05	-0.203542029
chr1	153517654	153517859	0.015459655	0.111432355
chr1	154956736	154957297	0.063435358	-0.088485589
chr1	160681478	160682169	0.037990596	-0.040921991
chr1	161038641	161039032	4.80E-09	-0.264158562
chr1	161475461	161475849	0.002652877	-0.16864877
chr1	173174911	173176087	0.082438786	-0.056016855
chr1	173449733	173449965	0.102115617	0.04274556
chr1	182357927	182358329	0.068397296	0.062977644
chr1	192544836	192545273	0.015727132	-0.133234131
chr1	198607773	198610160	0.018214222	-0.100339811
chr1	198651498	198651957	0.407309588	-0.037295152
chr1	199010586	199011304	0.176595525	-0.042360171
chr1	199660947	199661166	0.163636918	-0.04654987
chr1	211847053	211847783	0.01220632	0.100825287
chr1	212115502	212117111	0.205896102	0.032334064
chr1	220653057	220653280	0.173358467	0.062705846
chr1	220864886	220865396	0.191583115	-0.055482328

chr1	221909048	221909616	0.107649879	-0.033404002
chr1	230249386	230249800	0.067540877	-0.041459111
chr1	236144282	236144424	0.428590605	-0.012324001
chr1	241800452	241800722	0.004056696	-0.109795008
chr1	248021785	248022041	0.200965059	-0.040951299
chr2	10093863	10094145	0.149489146	-0.059018061
chr2	16084776	16085714	0.026462974	-0.037242163
chr2	21151188	21151541	0.006293461	-0.043007646
chr2	25668139	25668364	0.117496348	-0.099659161
chr2	38669779	38670055	0.192013286	0.04128462
chr2	44308569	44309760	0.183046934	0.088509005
chr2	45911195	45911602	0.277667197	-0.043119313
chr2	46795289	46796098	0.022247323	-0.09536594
chr2	47216781	47216960	0.001719588	0.178933354
chr2	62426656	62427018	0.021726827	-0.05800806
chr2	63270337	63271211	0.105911246	0.032484673
chr2	67625367	67625994	0.014192426	-0.159468397
chr2	68592241	68593052	0.000417575	-0.116575262
chr2	68961351	68961704	0.000287308	-0.169520485
chr2	68962521	68963097	0.000287308	-0.169520485
chr2	70211010	70211141	0.074793806	-0.040295727
chr2	80531572	80531693	0.01267564	-0.12761776
chr2	100758499	100758785	0.037745155	-0.103856175
chr2	120128320	120128605	0.096817539	0.042198961
chr2	120279902	120280158	0.049402728	-0.060074519
chr2	136870431	136870714	0.025691591	-0.101355347
chr2	145418645	145418943	0.013432772	-0.148539776
chr2	158298920	158300293	0.003845787	-0.126579923
chr2	158300943	158301783	0.003845787	-0.126579923
chr2	158759827	158760594	0.345419057	-0.037080746
chr2	177023147	177023471	0.00782395	-0.119911701
chr2	234160922	234161606	0.408179201	0.018866827
chr2	237078975	237079346	0.014302383	-0.228144487
chr2	241721037	241721263	0.157313624	0.04528784
chr6	1608907	1609124	0.02149946	-0.144758449
chr6	6701849	6702647	0.138073502	-0.044960274
chr6	8069397	8069892	0.230098757	0.03717546
chr6	10557112	10557387	0.084712725	0.085695391
chr6	20149392	20149620	0.357755126	-0.029006955
chr6	26595986	26596063	0.001774016	0.091203913
chr6	43488870	43489065	0.039655641	-0.063665591
chr6	43977929	43978138	0.067770138	-0.049733177
chr6	72001391	72001673	0.650063432	0.021514144

chr6	75913118	75913414	0.039964536	0.036258368
chr6	79922240	79922533	0.031223875	0.101608036
chr6	83774408	83774634	0.088148366	0.024298645
chr6	90995934	90996135	0.31033621	-0.018624689
chr6	106630448	106630669	0.032929012	-0.092864789
chr6	108145178	108145438	0.0002544	-0.15054077
chr6	109629070	109629701	0.07455092	-0.113810962
chr6	121655554	121655699	0.214211871	-0.076381082
chr6	122721624	122721938	0.782800905	-0.005988991
chr6	135946165	135946429	0.150413734	-0.027839173
chr6	136376869	136377151	0.138900626	0.033768848
chr6	139692242	139692766	0.071358568	-0.05626553
chr6	142683092	142683312	0.060847098	0.063670558
chr6	154567738	154568040	0.054281324	-0.064447241
chr6	163814936	163815662	0.005485027	0.144576733
chr6	164542172	164542650	0.06111785	-0.049408781
chr8	1715132	1715419	0.020312359	-0.055086638
chr8	17529635	17529755	0.212448998	-0.045221446
chr8	23184678	23184867	0.033107474	-0.028156664
chr8	26429564	26429693	0.089238216	-0.060735815
chr8	28226653	28226969	0.070140437	-0.097469136
chr8	29191531	29191754	0.074664184	0.033243325
chr8	30402021	30402307	0.059656602	0.066092099
chr8	37384529	37384565	0.056355149	-0.074172948
chr8	41907137	41907600	0.033951722	-0.049325479
chr8	43152678	43152944	0.101774677	0.10468965
chr8	48921440	48921670	0.549728625	0.045360426
chr8	66701734	66702089	0.000239452	0.133829622
chr8	70588403	70588535	0.03737239	-0.118897956
chr8	72271032	72271813	0.006526032	0.055557647
chr8	79577640	79577971	0.027335434	-0.120166491
chr8	86377077	86377304	0.205873288	0.079580679
chr8	110346195	110346407	0.005606233	-0.146137424
chr8	124050010	124051132	0.016160863	-0.201700733
chr8	125642847	125643100	0.053762046	0.044152252
chr8	125671231	125671900	0.105146153	0.033045224
chr8	126011014	126011422	0.018217776	0.088449027
chr8	128941581	128942080	0.029892253	0.074853732
chr8	130956680	130957583	0.006432972	0.083997065
chr8	131457070	131457446	0.091831138	-0.05001441
chr8	141173973	141174525	0.457540782	0.018224753
chr8	144816484	144816861	0.026970252	-0.06895942
chr10	1402817	1403101	0.008030471	-0.051764156

chr10	3895156	3895674	0.055394379	0.036857931
chr10	5485881	5486200	0.081087469	-0.097455401
chr10	5490224	5490845	0.149235639	0.058475426
chr10	5708419	5708659	0.059598542	-0.04917547
chr10	6190543	6191242	0.017146089	-0.073584795
chr10	21571119	21571557	0.30595094	-0.036460642
chr10	25305780	25305987	0.101278552	-0.053808868
chr10	70884710	70884845	0.087081301	-0.069417137
chr10	73300885	73301322	0.080482241	0.055621735
chr10	82175262	82175938	0.130456278	0.059308364
chr10	82180528	82180732	0.061117646	0.080669928
chr10	82279570	82279941	0.013820019	-0.139269935
chr10	91006627	91007328	0.074112974	-0.077302683
chr10	97144444	97145224	0.027219413	0.07734614
chr10	97517128	97517433	0.026474464	-0.077219508
chr10	98416374	98417030	0.002273819	-0.160411722
chr10	99160683	99160877	0.080496726	-0.057407697
chr10	102109210	102109419	0.049406421	-0.040765323
chr10	103245631	103246048	0.057694353	-0.085267828
chr10	104632622	104633383	0.004186243	0.118357444
chr10	105414950	105415243	0.000643737	0.11508831
chr10	112839077	112839216	0.153991099	-0.061788635
chr10	114760114	114760561	0.00035593	0.158699963
chr10	114801196	114802025	0.001798348	0.127830599
chr10	114833951	114834429	0.100312432	0.057000362
chr10	115824088	115824893	0.272925161	-0.0481567
chr10	117577292	117577586	0.162784849	0.055245791
chr10	122610736	122611078	0.157144273	-0.045241863
chr10	123920852	123920995	0.042701586	-0.050719844
chr10	124135396	124135700	0.198129182	0.050715394
chr10	128592925	128593141	0.055414321	-0.036175689
chr10	129846102	129846548	0.006558867	-0.157671746
chr10	132127617	132128103	0.259498246	0.021822295
chr10	132897979	132898573	0.023710989	0.071039385
chr10	133793796	133793957	0.032971233	0.105305104
chr10	135072811	135073045	0.04716606	-0.118765257
chr10	135349833	135351027	0.098711704	0.049678185
chr12	1663136	1663829	0.043716238	-0.046192522
chr12	6166239	6166263	0.248521563	0.043982022
chr12	6469500	6469539	0.03543076	0.03100671
chr12	7060921	7061200	0.000293744	-0.145500528
chr12	7308164	7308415	0.035942937	-0.126870767
chr12	8216915	8217820	0.114586135	-0.033344046

chr12	8275965	8276317	0.013803864	-0.124748032
chr12	11812172	11812279	0.005150697	0.120887626
chr12	14928365	14928608	0.115301893	-0.022948594
chr12	14995468	14996149	0.047220256	0.076344607
chr12	15111717	15112453	0.00956376	-0.100766318
chr12	26392741	26393036	0.032865677	0.114081438
chr12	40010029	40010986	0.045439035	-0.052089161
chr12	50419685	50419898	0.067364544	-0.028632215
chr12	53268780	53268925	0.417767017	-0.0228861
chr12	53497808	53497996	0.059851859	0.04586622
chr12	54686256	54686746	0.047686589	-0.052394071
chr12	62658334	62658491	0.012820313	-0.080748795
chr12	69140644	69140871	0.00294898	-0.05840808
chr12	75874872	75875171	0.028932969	-0.079091903
chr12	76662008	76662319	0.230407077	0.084261368
chr12	94978299	94978583	0.117548072	-0.021978496
chr12	95000081	95001986	0.004387317	0.116714272
chr12	96813903	96814310	0.097873514	0.084093689
chr12	100538575	100538610	0.2054361	0.045371194
chr12	101673825	101674109	0.068941654	-0.057969332
chr12	102093732	102095643	0.086619668	0.103113517
chr12	102618254	102618882	0.004350063	0.115790802
chr12	104363097	104363317	0.005057558	-0.136091463
chr12	106621844	106621970	0.036423307	0.046232273
chr12	107045994	107046229	0.452216789	0.039977643
chr12	109194765	109194928	0.017916085	0.035229392
chr12	109838899	109839351	0.495308113	0.036933183
chr12	110436172	110436489	0.240157029	0.050905326
chr12	115108903	115109193	0.039598932	-0.062105131
chr12	115125816	115126012	0.00275724	-0.14756063
chr12	115129093	115129959	0.00030568	-0.187331894
chr12	117173023	117173043	0.095202773	-0.072915207
chr12	117433133	117434138	0.047336201	-0.0429966
chr12	118796639	118796981	0.030025023	-0.061604516
chr12	122204046	122204675	0.013695007	-0.095305859
chr12	124865362	124865831	0.102138472	0.049625685
chr12	125333888	125334033	0.171709363	0.077782925
chr12	125575960	125576193	0.115106488	0.074375332
chr22	18298189	18298467	0.036134946	0.088448521
chr22	24761416	24761463	0.344188283	0.014164191
chr22	30278371	30278842	0.162418389	0.047787343
chr22	36924935	36925290	0.033885498	-0.073884507
chr22	38472613	38473870	0.03773117	-0.050802924

chr22	40858582	40860665	0.001384889	-0.125891594
chr22	43555410	43557147	0.018177759	-0.050771113
chr22	45561031	45561107	0.264933651	0.033646427
chr22	50955702	50956285	0.069548977	0.071080994
chr17	1220102	1220359	0.00498105	0.124949686
chr17	7483300	7483645	0.004383267	-0.086940803
chr17	8027102	8027420	0.039741912	-0.033411302
chr17	8867824	8868379	0.128040105	-0.057670787
chr17	9950908	9951126	0.208451228	-0.047643312
chr17	26646057	26646266	0.009588159	-0.219201828
chr17	29150010	29150903	0.064732777	-0.026243112
chr17	29641021	29641319	0.137955415	-0.060964811
chr17	34415897	34416118	0.001156209	-0.090858161
chr17	34417557	34418034	0.001156209	-0.090858161
chr17	35299693	35299863	0.353425615	0.019486541
chr17	40142289	40142472	0.07140143	0.050557861
chr17	45908521	45908890	0.158883651	0.046498288
chr17	47819090	47819795	0.007399875	-0.102310745
chr17	47863126	47864532	0.071136849	0.073246062
chr17	48940751	48941296	0.06118011	0.085294662
chr17	56124678	56125091	0.022776363	-0.076070557
chr17	56355449	56355594	0.038596305	-0.060609612
chr17	56490498	56491520	0.022309925	0.109264262
chr17	62204776	62206576	0.076370686	-0.096781936
chr17	63541217	63541473	0.020481163	-0.049730423
chr17	64448688	64449382	0.14072088	-0.047904076
chr17	66877696	66877766	0.122227015	-0.086695187
chr17	73406576	73407600	0.110354272	-0.048225572
chr17	73630171	73630823	0.006844407	-0.104660956
chr17	76375529	76375618	0.160821825	0.071059014
chr17	79419543	79420749	0.017498542	0.105982501
chr17	80050393	80050984	0.035750412	-0.102212516
chr17	80052124	80053059	0.00234477	0.110938993
chr17	80416305	80416449	0.377968144	0.039260145
chr5	10304741	10305342	0.016968763	-0.102322221
chr5	32780268	32780484	0.043125176	-0.062892296
chr5	34655547	34655935	0.04433336	0.097710785
chr5	34687673	34687965	0.027202557	0.113639208
chr5	36535659	36536198	0.188674087	-0.035795174
chr5	55737286	55737497	0.016894229	0.094145368
chr5	64333579	64333820	0.011515023	0.080503852
chr5	66491448	66492361	0.010515394	-0.065840584
chr5	67585536	67585875	0.027604432	0.069279842

chr5	74244430	74245271	0.051883314	-0.077775325
chr5	77803248	77803367	0.014339375	-0.094089602
chr5	79028795	79029354	0.404746582	-0.014708512
chr5	106908226	106908482	0.040259849	0.089666375
chr5	106938858	106939295	0.103233424	0.063036468
chr5	108085570	108085868	0.014590337	-0.144368703
chr5	133968325	133968612	0.19291661	0.155371226
chr5	134787566	134787794	0.121578462	-0.04209435
chr5	141670403	141670927	0.093448335	0.087272102
chr5	148208221	148208626	0.005235308	-0.099159277
chr5	149108648	149108898	0.070181892	0.0958248
chr5	149464925	149465615	0.008300488	-0.087832678
chr5	150726558	150727030	0.333779062	0.04041369
chr5	169760638	169760900	1.42E-05	-0.194462311
chr5	176325586	176326273	0.178871723	-0.050743362
chr5	178803895	178804058	0.010600527	-0.111699146
chr7	1408433	1408951	0.01030415	0.063819595
chr7	8302335	8302647	0.852262755	0.008577271
chr7	26460385	26460633	0.017227015	-0.112177321
chr7	27201206	27201499	0.000244	-0.154545194
chr7	27202553	27202993	0.00772616	-0.163900092
chr7	27203073	27203466	0.00772616	-0.163900092
chr7	30287772	30288094	0.502192871	0.028442713
chr7	47672033	47673452	0.001650023	0.090694865
chr7	50345715	50346835	0.000203584	-0.123971624
chr7	73624879	73625346	0.025835427	-0.095097771
chr7	75549990	75550275	0.240896483	-0.019749676
chr7	75957929	75958231	0.014476618	-0.079952002
chr7	76828348	76828576	0.032838026	-0.069069628
chr7	97500452	97501026	0.05303718	0.096679458
chr7	100238885	100239167	0.074148968	0.074548168
chr7	105318767	105318983	0.001087181	0.126246501
chr7	106704715	106705240	0.001244862	0.088189068
chr7	107609751	107610536	0.256006703	0.064722568
chr7	107692799	107693160	0.020384645	-0.090978916
chr7	127857315	127858302	0.193642885	0.043408718
chr7	127889822	127890030	0.177358186	0.052936735
chr7	130622635	130623162	0.062198878	0.062047453
chr7	131241745	131242050	0.099786111	-0.055238515
chr7	139529064	139530436	7.05E-05	-0.203984492
chr7	140048085	140048549	0.219770505	0.023951494
chr7	151039508	151039684	0.00124038	0.194317077
chr14	20898926	20899114	0.001031265	-0.133171333

chr14	21568858	21569390	0.227490337	0.046680964
chr14	23788798	23789618	0.113051505	-0.098799804
chr14	31888216	31888303	0.051428286	-0.077015105
chr14	39737165	39737329	0.09171178	0.047880934
chr14	55833724	55834573	0.128314832	0.019852413
chr14	55834889	55835484	0.128314832	0.019852413
chr14	74961318	74961491	0.02971491	-0.068992065
chr14	75361059	75361571	0.002412449	-0.165986778
chr14	77791405	77791671	0.010968896	0.134177876
chr14	81867674	81868031	0.081282534	-0.073907314
chr14	89707672	89708616	0.005943992	0.082570716
chr14	93117257	93117768	0.003786645	-0.112199089
chr14	97180040	97180391	0.161683822	0.025082244
chr14	100531888	100532636	0.003391464	-0.113839928
chr9	14203697	14204081	0.001304379	0.122130904
chr9	32457089	32457364	0.019332685	0.140039676
chr9	36778847	36778989	0.017411606	-0.117459817
chr9	93565879	93566460	0.007420128	0.094453504
chr9	95726604	95726801	0.000782954	-0.190154032
chr9	113021805	113022191	0.153227162	0.059319005
chr9	113310163	113310641	0.635801168	-0.026481336
chr9	114424532	114424918	0.008616226	-0.054665936
chr9	116379329	116379646	0.034041424	0.053038017
chr9	117085126	117085447	0.089241078	0.033000789
chr9	117147509	117147976	0.090278173	0.057770456
chr9	117692618	117692950	0.000933629	-0.161745875
chr9	130281483	130281851	0.121434378	-0.03946649
chr9	131369913	131370522	0.021422761	0.053623327
chr9	132804046	132804253	0.060174876	-0.083590765
chr13	29293918	29294030	0.234074293	0.075604237
chr13	32780792	32781098	0.477862977	0.019990114
chr13	36049213	36050887	0.000513404	-0.083527653
chr13	40809925	40810619	0.449595212	0.027281116
chr13	40943529	40943850	0.045248093	0.073827692
chr13	44936911	44937231	0.071609449	-0.039506342
chr13	45564095	45564224	0.004529161	-0.179434382
chr13	45694501	45694858	0.004084029	-0.071848765
chr13	46751279	46752608	0.014544051	-0.091356551
chr13	46753288	46754147	0.014544051	-0.091356551
chr13	49683338	49683473	0.149221767	0.029842321
chr13	50204794	50205145	0.044279538	-0.107739618
chr13	114018968	114019138	0.229393431	-0.053562829
chr3	9911613	9911971	0.014015847	0.066336204

chr3	15667550	15667785	0.057143936	-0.043614963
chr3	15681784	15682347	0.152744834	-0.05127628
chr3	24869934	24870287	0.029777881	0.046387087
chr3	32412954	32413251	0.172866571	0.03624256
chr3	37907270	37907753	0.374453931	0.023449625
chr3	38007994	38008396	0.031627285	-0.085747522
chr3	40499457	40499546	0.026209488	0.061704304
chr3	46036856	46037289	0.013540094	-0.234110654
chr3	52445047	52445468	0.164356522	0.045244721
chr3	52567262	52567405	0.130156467	-0.037378327
chr3	57262305	57262906	0.239989212	-0.051311872
chr3	65434405	65434559	0.049562076	-0.048552937
chr3	66633695	66634011	0.087614339	-0.038846936
chr3	67049442	67049635	0.159046189	0.057621246
chr3	72870608	72870903	0.026213972	-0.070105607
chr3	72896699	72897027	0.004251486	-0.192497184
chr3	112692217	112693342	0.009566716	-0.067486844
chr3	112964544	112965111	0.026237876	0.087180801
chr3	124349805	124350028	0.053370691	-0.053070703
chr3	124511021	124511255	0.112577227	0.052593515
chr3	124511634	124511961	0.112577227	0.052593515
chr3	124559262	124559467	0.33380135	0.051607354
chr3	127512341	127512749	0.300572926	0.046883681
chr3	133615194	133615551	0.012336117	0.053271596
chr3	138311777	138311800	0.091346454	0.055603918
chr3	141108490	141109631	0.01672593	-0.121198507
chr3	150322081	150322440	0.000503571	-0.14397228
chr3	150965754	150966036	0.115599	-0.047542803
chr3	151102250	151102484	0.314598291	0.057538343
chr3	157325748	157326168	0.000152002	-0.188356272
chr3	159706955	159707276	0.510455058	-0.037543221
chr3	169897652	169897802	0.06071042	-0.081619325
chr3	170894740	170894972	0.246755268	0.062234751
chr3	185226120	185226426	0.11891308	-0.025769524
chr3	186500542	186500813	0.057501757	0.053855992
chr3	187437450	187437801	0.183180484	-0.023522408
chr3	188585553	188586208	0.255225767	0.039438945
chr3	196063662	196064900	0.200280698	-0.077940287
chr3	196304666	196305922	0.051206131	0.032303493
chr16	16123803	16124667	0.028137766	0.062463282
chr16	20909935	20910137	0.085974868	0.056191121
chr16	30455438	30455975	0.03253984	-0.0577708
chr16	46957730	46958679	0.031950761	-0.034656493

chr16	50059655	50059880	0.063328638	-0.103644757
chr16	50715116	50715299	0.000204672	-0.169680371
chr16	55540360	55540833	0.012314177	-0.055940936
chr16	56137894	56138256	0.004769343	0.066758804
chr16	57278607	57278995	0.317131448	0.045084544
chr16	69457347	69457490	0.01244399	0.086794906
chr16	79370715	79370926	0.118336141	0.035616683
chr16	83549673	83549934	0.092723296	0.034427003
chr16	85376056	85376680	0.182709459	-0.023703324
chr16	87903345	87904117	0.034324129	-0.089302202
chr16	88770501	88770848	0.043240371	-0.073898435
chr21	16057979	16058284	0.414561669	-0.033457948
chr21	33689979	33690246	0.149919203	-0.039646958
chr21	35448249	35449067	0.003752114	-0.121181521
chr21	36398548	36398828	0.000151796	-0.15178274
chr21	36420456	36421805	4.31E-06	-0.279855942
chr21	37412563	37413036	0.057139401	0.073919244
chr21	39868683	39869555	0.000107641	-0.134357154
chr21	40776034	40777497	0.304302694	0.020796314
chr18	2960647	2960946	0.226393124	0.074511207
chr18	10457031	10457507	0.093670381	0.042061125
chr18	46277273	46277494	0.060855394	0.080607141
chr18	56809110	56809370	0.047576474	-0.102872363
chr18	77440474	77440960	0.124164823	-0.03050826
chr19	7733678	7734230	0.017884245	-0.067224475
chr19	12920962	12921262	0.070335289	-0.037572965
chr19	15574480	15574831	0.019917347	-0.119050573
chr19	17669572	17669647	0.503151826	-0.027583264
chr19	18508116	18508457	0.155971856	-0.0272848
chr19	18682314	18682390	0.137872609	0.056746254
chr19	40750567	40750963	0.00536166	0.105072993
chr19	42830523	42830820	0.017467435	-0.125098373
chr19	42912089	42912557	0.063280265	-0.066215746
chr19	45225646	45225969	0.096953319	0.111422729
chr19	49841249	49841895	0.027967301	0.091393575
chr19	51645546	51714605	0.026484673	-0.072463543
chr19	54876429	54876765	0.026489679	-0.097348731
chr11	6267881	6268720	0.001427368	0.142949402
chr11	10316893	10317243	0.015285684	-0.079842283
chr11	10324978	10325218	0.006833463	-0.076195352
chr11	13359375	13359540	0.287397995	0.033110364
chr11	14563296	14563419	0.268961707	0.042292388
chr11	18127092	18127637	0.22057406	0.022446951

chr11	27490752	27491523	0.020887869	-0.168418694
chr11	32852321	32852814	0.181458128	0.058426216
chr11	47291136	47291346	0.117725556	-0.06055027
chr11	47415025	47415403	0.005612808	-0.09283909
chr11	57067776	57068975	0.031735717	-0.108159774
chr11	63528336	63528808	0.009374816	0.075120856
chr11	64544695	64545147	0.074007346	0.080573957
chr11	66673568	66673798	0.049344564	0.089939048
chr11	67036383	67036711	0.000861836	-0.127008445
chr11	67165584	67166226	0.038535186	0.032864959
chr11	73692113	73692198	0.048032778	-0.086506376
chr11	76337568	76338271	0.052017168	-0.077393197
chr11	117858523	117858767	0.000608402	-0.109964715
chr11	120987491	120987707	0.070233309	-0.032303992
chr11	122849412	122849982	0.034711309	0.095100232
chr15	37173455	37174113	0.028601312	-0.075724131
chr15	52404555	52404986	0.115604037	0.071796217
chr15	52820075	52820295	0.694329386	0.024475175
chr15	61136559	61136807	0.001200488	-0.123028637
chr15	62343394	62344034	0.214289764	-0.064912798
chr15	63571085	63571484	0.284299059	-0.051881135
chr15	69426262	69426542	0.018055549	-0.112554107
chr15	69708054	69708318	0.020715438	0.049924329
chr15	77478505	77479038	0.231666602	0.070100851
chr15	79471213	79471368	0.074622439	-0.035916955
chr15	89977287	89977470	0.00877423	-0.073352554
chr15	90701174	90701775	0.085843566	0.0589646
chr15	96902600	96903113	4.02E-05	-0.13012121
chr15	96905568	96906197	0.000917372	-0.151366539
chr15	96907528	96907914	0.000917372	-0.151366539
chr20	1879865	1880324	0.001414611	0.163898814
chr20	3527981	3528426	0.147326039	0.084492638
chr20	3746584	3746692	0.139490039	-0.044307576
chr20	18269464	18269677	0.004138053	0.11442861
chr20	25033629	25034262	0.035246291	0.053724766
chr20	25251494	25252151	0.00912653	0.050518875
chr20	30258398	30258882	0.062783463	0.088332976
chr20	33466246	33466396	0.135099492	0.049631059
chr20	35233479	35233726	0.178100009	-0.054084787
chr20	36104587	36105045	0.155524754	-0.044843164
chr20	36823215	36823538	0.060042606	-0.026596859
chr20	39314354	39315324	0.016335516	-0.073400089
chr20	43953416	43953593	0.222068845	-0.027653814

chr20	47803610	47804113	0.04857818	-0.099555153
chr20	47839439	47839685	0.156286748	-0.02991861
chr20	48924232	48924768	0.004243497	0.096121482
chr20	56135219	56135721	0.008351783	0.073282322
chr20	56135935	56137320	0.003187762	0.087199606
chr20	56138648	56138937	0.003187762	0.087199606
chr20	56139464	56139982	0.003187762	0.087199606
chr20	58512538	58512948	0.099138641	0.056386238
chr4	3295253	3295496	0.21813933	0.078202348
chr4	38858508	38859749	0.003242039	-0.125613648
chr4	39408228	39410160	0.044594542	-0.072267154
chr4	84034389	84034401	0.000415556	-0.184889567
chr4	86398976	86399113	0.063282407	0.060784671
chr4	87672032	87672388	0.100709592	-0.079778073
chr4	108552737	108553088	0.052559898	-0.072581024
chr4	108928564	108928770	0.04453389	0.055177969
chr4	113861384	113862156	0.027308541	-0.078932185
chr4	119855498	119855703	0.1507209	0.081634037
chr4	120547931	120548111	0.051707622	0.108026119
chr4	129162516	129162816	0.046803004	-0.037213593
chr4	129230121	129231987	0.145876681	0.034008826
chr4	184241311	184241863	0.04719212	0.079883969
chr4	184774713	184775223	0.010583934	0.096433022
chr4	185303774	185304147	0.044240008	-0.068515138
chr4	185741474	185741616	0.21397322	0.051362717
chr4	185743539	185744674	0.207983162	-0.070745646
chr4	185796824	185797589	0.174978393	0.035307295

Table 7**. Conserved mouse-human DMRs.** This table lists the 497 mouse DMRs mappable onto the human chromosome and with 5kb of a human probe. Listed are the genomic coordinates and width for each mouse differentially methylated region (DMR), q-values for the mouse DMRs derived from false discovery rate (see methods, qval), the gene symbol nearest gene to the mouse DMR, the p-values for the corresponding changes in human obesity and surgery, and the slopes for the methylation change for both human obesity and surgery.

Table 8

Table 8**. Enrichment of cross-species DMRs over DIAGRAM GWAS loci**

DIAGRAM		Number of GWAS loci overlapping conserved DMRs		Number of GWAS loci overlapping conserved, same direction DMRs	
Cutoff	# Loci	Adipose	Islet	Adipose	Islet
p<5e-08	18	**2 (0.017)**	0 (0.292)	**1 (0.0113)**	0 (0.108)
p<1e-07	20	**2 (0.025)**	0 (0.329)	**1 (0.0154)**	0 (0.121)
p<1e-06	35	**3 (0.016)**	1 (0.119)	**1 (0.0376)**	0 (0.179)
p<1e-05	215	**6 (0.005)**	2 (0.119)	**2 (0.0253)**	1 (0.058)
p<1e-04	428	9 (0.209)	5 (0.239)	3 (0.192)	2 (0.166)
p<1e-03	1601	24 (0.48)	12 (0.600)	6 (0.597)	7 (0.086)
p<1e-02	8353	77 (0.336)	39 (0.560)	27 (0.119)	12 (0.573)

Table 8**. Enrichment of cross-species DMRs over DIAGRAM GWAS loci.** This table summarizes the number and significance of overlaps of cross-species conserved adipose and pancreatic islet loci with DIAGRAM GWAS LD-blocks associated with SNPs at varying levels of significance as indicated by the cutoff column.

Table 9

Table 9**: Conserved mouse-human DMRs with genetic T2D risk loci association**

Nearest gene	Location relative to gene	Distance to TSS	RYGB Reversion	DIAGRAM P-value	Diabetes annotation
Tcf7l2	inside intron	43058	-	4.90E-68	Strongest known genetic association with T2D (Morris et al., 2012)
Tcf7l2	inside intron	77345	-	4.90E-68	Strongest known genetic association with T2D (Morris et al., 2012)
As3mt	overlaps 5'	0	+	9.60E-06	Non-genome wide association with T2D (Drobna et al., 2013)
Etaa1	inside intron	618	+	4.70E-05	Suggestive genetic association with waist-hip ratio (Liu et al., 2013a)
Tnfsf8	overlaps 5'	0	-	0.00029	Known cytokine; important role in development of Type 1 Diabetes in mice (Chakrabarty et al., 2003)
Plekho1	overlaps exon	4965	+	0.00045	**Tested functionally**
Tnfaip8l2	inside intron	337	+	0.00045	**Tested functionally**
Akt2	inside intron	20427	-	0.00049	Central component in insulin-signaling pathway (Cho et al., 2001)
DIAGRAM GWAS 0.001 cutoff					
Lhfpl2	inside intron	2490	+	0.001	
Mkl1	overlaps 5'	0	+	0.0014	**Tested functionally**
BC048644 (overlaps Car5a)	overlaps exon	146	+	0.0015	Putative mouse transcript; overlaps with the *Car5a* gene **Tested functionally**

Rgs3	downstream	108842	+	0.0019	
Fgd3	inside intron	11100	+	0.002	Differentially expressed in animal models of impaired insulin signaling (Zarse et al., 2012)
Stau1	overlaps 5'	0	+	0.0022	Impaired response to insulin in muscle in T2D subjects (Rome et al., 2009)
Tmcc3	inside intron	43772	+	0.0025	**Tested functionally**
Tbx3	inside exon	12714	-	0.0029	
Gstz1	inside intron	10332	+	0.0029	**Tested functionally**
Taok3	inside intron	549	+	0.0036	
Bnip3	inside intron	1863	-	0.0039	Contributes to oxidative stress in mouse model diabetes (Su et al., 2008)
Dlst	overlaps 5'	0	+	0.0053	
Kcna3	close to 3'	2192	+	0.0064	Regulates peripheral insulin sensitivity (Xu et al., 2004)
Cln8	inside intron	3055	+	0.0065	
Cd37	overlaps exon	2687	+	0.0069	
Nfib	inside intron	100380	-	0.0071	Impaired response to insulin in muscle in T2D subjects (Rome et al., 2009)
Pck1	promoter	453	+	0.0072	Central component in de novo lipogenesis and insulin resistance (Millward et al., 2010)
Pck1	overlaps 5'	0	+	0.0072	Central component in de novo lipogenesis and insulin resistance (Millward et al., 2010)
Pcx	inside intron	59049	+	0.0073	Inhibition of pyruvate carboxylase inhibits gluconeogenesis (Bahl et al., 1997)
Hoxd3	inside intron	7307	+	0.0084	Accelerates wound healing in diabetic mice (Hansen et al., 2003)

Cd33	overlaps 5'	0	=	0.0087	Significantly decreased expression in monocytes from T2D subjects (Gonzalez et al., 2012)
Evl	covers exon(s)	157	+	0.0099	

BLUE – Genes previously found by conventional GWAS as being associated with T2D after multiple testing correction. RED – Genes near conserved, cross-species epigenetic changes that overlap LD blocks where genetic variants have an individually significant association with T2D. Shown are the names of the nearest gene to the mouse and human differential methylation, the position of the DMR relative to the gene (only one DMR is outside the gene), the distance to the transcriptional start site (TSS), whether the direction of methylation change (sign of smoothed effect statistic) post-RYGB surgery reverts toward lean subject methylation levels (RYGB Reversion), and the p-value of the T2D genetic association in the region.

Table 10

Table 10**. Cross-species, directionally consistent pancreatic islet DMRs that overlap with DIAGRAM T2D GWAS loci**

Mouse DMR Chromosome	Mouse DMR start	Mouse DMR end	Mouse DMR width	Mouse DMR value
chr3	97536603	97536878	276	0.093610895
chr1	147083046	147084059	1014	0.079418323
chr2	169757423	169758143	721	0.08021799
chr11	17405939	17406583	645	0.080153144
chr11	17406741	17407082	342	0.077431401
chr11	17407373	17408268	896	0.081308921
chr11	27491277	27491838	562	-0.078049539
chr14	77511555	77511688	134	-0.081407373
chr12	39297915	39298205	291	-0.071278015
chr5	123149655	123150023	369	0.073113986

pval	qval	fwer	nearestGene	nearestGeneDesc	nearestGeneDist
0.0015625	0.012831439	0.018	Arl6	downstream	34177
6.51E-05	0.002027957	0	Bcl9	overlaps 5'	0
6.51E-05	0.002027957	0	G6pc2	overlaps 5'	0
0.001497396	0.012831439	0.017	Kcnj11	downstream	3320
0.002929688	0.017074093	0.033	Kcnj11	overlaps 3'	2900
6.51E-05	0.002027957	0	Kcnj11	inside	1740
0.000911458	0.010692866	0.011	Lgr4	inside	2417
0.016080729	0.045394558	0.123	6430527G18Rik	upstream	15082
0.004231771	0.020119472	0.044	Cpne8	inside	926
0.001953125	0.013892306	0.021	Csnk1g3	downstream	263645

HumanPval (for T2D)	DIAGRAM p-value of nearest GWAS loci	DIAGRAM SNP
0.018109593	0.00069	rs17302349
0.018785468	0.00042	rs7512513
0.005552449	0.0034	rs16856159
0.011107651	4.40E-06	rs5215
0.011107651	4.40E-06	rs5215
0.001281212	4.40E-06	rs5215
0.045056138	0.0077	rs7945211
0.024935476	0.0029	rs17752640
0.008167247	1.60E-05	rs10506132
0.010390504	0.00051	rs7708937

Table 10**. Cross-species, directionally consistent DMRs that overlap with DIAGRAM T2D GWAS loci.** Similar to Table 3, this table lists pancreatic islet DMRs that are significant across species, directionally consistent, and overlap with DIAGRAM T2D LD blocks associated with nominally significant SNPs.

Table 11

Table 11. **Overlap of conserved and directionally consistent adipose DMRs and adipose enhancers**					
chr	start	end	Mouse_Nearest_Gene	distToNearestAdiposeEnhancer	coordsOfNearestAdiposeEnhancer
chr1	19804495	19805076	Capzb	0	chr1:19739675-19833107
chr1	48351513	48351750	OTTMUSG00000008561	-546242	chr1:47798033-47805271
chr1	64061530	64062519	Pgm2	0	chr1:64055024-64092549
chr1	111214849	1.11E+08	Kcna3	199363	chr1:111414700-111418046
chr1	114413127	1.14E+08	Ptpn22	-13238	chr1:114398212-114399889
chr1	114414346	1.14E+08	Ptpn22	-14457	chr1:114398212-114399889
chr1	114415254	1.14E+08	Ptpn22	-15365	chr1:114398212-114399889
chr1	150129607	1.5E+08	Plekho1	0	chr1:150113475-150160132
chr1	151129480	1.51E+08	Tnfaip8l2	-10710	chr1:151103285-151118770
chr1	160681478	1.61E+08	Cd48	-131593	chr1:160543677-160549885
chr1	161038641	1.61E+08	Arhgap30	0	chr1:161037436-161039509
chr1	161475461	1.61E+08	Fcgr3	23356	chr1:161499205-161533394
chr1	192544836	1.93E+08	Rgs1	-14758	chr1:192528534-192530078
chr1	198607773	1.99E+08	Ptprc	16327	chr1:198626487-198627845
chr1	211847053	2.12E+08	Nek2	-55976	chr1:211784613-211791077
chr2	16084776	16085714	Mycn	-238915	chr2:15843316-15845861
chr2	47216781	47216960	Ttc7	0	chr2:47166289-47314830
chr2	62426656	62427018	B3gnt2	-588	chr2:62421832-62426068
chr2	67625367	67625994	Etaa1	-136215	chr2:67486284-67489152
chr2	68592241	68593052	Plek	0	chr2:68592492-68596426
chr2	68961351	68961704	Arhgap25	28829	chr2:68990533-69007401
chr2	68962521	68963097	Arhgap25	27436	chr2:68990533-69007401
chr2	100758499	1.01E+08	Lonrf2	49822	chr2:100808607-100816096
chr2	145418645	1.45E+08	Zeb2	0	chr2:145417385-145430018
chr2	158298920	1.58E+08	Pscdbp	53777	chr2:158354070-158355373
chr2	158300943	1.58E+08	Pscdbp	52287	chr2:158354070-158355373
chr2	177023147	1.77E+08	Hoxd3	-17061	chr2:176992872-177006086
chr6	1608907	1609124	Foxc1	2698	chr6:1611822-1614843
chr6	26595986	26596063	Abt1	-22946	chr6:26571587-26573040
chr6	75913118	75913414	Col12a1	0	chr6:75909506-75944578
chr6	79922240	79922533	Hmgn3	19305	chr6:79941838-79944997
chr6	108145178	1.08E+08	Scml4	128176	chr6:108273614-108279958
chr6	163814936	1.64E+08	Qk	0	chr6:163812425-163859943
chr8	1715132	1715419	Cln8	0	chr8:1693029-1737972
chr8	41907137	41907600	Myst3	0	chr8:41890037-41909112
chr8	66701734	66702089	Pde7a	50510	chr8:66752599-66755870
chr8	70588403	70588535	Slco5a1	-70101	chr8:70516915-70518302
chr8	79577640	79577971	3110050N22Rik	-49109	chr8:79527654-79528531
chr8	128941581	1.29E+08	Myc	-1615	chr8:128938286-128939966
chr8	130956680	1.31E+08	0910001A06Rik	0	chr8:130944190-131028814

chr8	144816484	1.45E+08	AA409316	-100509	chr8:144714373-144715975
chr10	6190543	6191242	Rbm17	0	chr10:6182494-6261722
chr10	97144444	97145224	Sorbs1	0	chr10:97067610-97208276
chr10	97517128	97517433	Entpd1	-14493	chr10:97501217-97502635
chr10	98416374	98417030	Pik3ap1	15807	chr10:98432837-98434307
chr10	104632622	1.05E+08	As3mt	-53262	chr10:104497945-104579360
chr10	105414950	1.05E+08	Sh3pxd2a	0	chr10:105396542-105439328
chr10	114760114	1.15E+08	Tcf7l2	0	chr10:114708122-114899787
chr10	114801196	1.15E+08	Tcf7l2	0	chr10:114708122-114899787
chr10	129846102	1.3E+08	Ptpre	0	chr10:129845340-129863251
chr10	133793796	1.34E+08	Bnip3	0	chr10:133785164-133796990
chr10	135072811	1.35E+08	6430531B16Rik	0	chr10:135072055-135075460
chr12	1663136	1663829	Wnt5b	21029	chr12:1684858-1721862
chr12	7060921	7061200	Ptpn6	501	chr12:7061701-7063206
chr12	8275965	8276317	Clec4a4	32585	chr12:8308902-8310793
chr12	15111717	15112453	Arhgdib	0	chr12:15101053-15116243
chr12	26392741	26393036	Sspn	0	chr12:26391495-26392980
chr12	62658334	62658491	Usp15	272793	chr12:62931284-62932711
chr12	69140644	69140871	Slc35e3	-62479	chr12:69058087-69078165
chr12	75874872	75875171	Glipr1	0	chr12:75873225-75877983
chr12	95000081	95001986	Tmcc3	6592	chr12:95008578-95017711
chr12	102618254	1.03E+08	Pmch	-49193	chr12:102565262-102569061
chr12	104363097	1.04E+08	Tdg	-1480	chr12:104357862-104361617
chr12	115108903	1.15E+08	Tbx3	0	chr12:115106745-115123443
chr12	115125816	1.15E+08	Tbx3	-2373	chr12:115106745-115123443
chr12	115129093	1.15E+08	Tbx3	-5650	chr12:115106745-115123443
chr12	118796639	1.19E+08	Taok3	0	chr12:118794373-118796651
chr22	40858582	40860665	Mkl1	-17872	chr22:40839368-40840710
chr17	1220102	1220359	Tusc5	0	chr17:1170362-1234434
chr17	7483300	7483645	Cd68	0	chr17:7474748-7488471
chr17	8027102	8027420	Hes7	2176	chr17:8029596-8030860
chr17	34415897	34416118	Ccl3	-114527	chr17:34299542-34301370
chr17	34417557	34418034	Ccl3	-116187	chr17:34299542-34301370
chr17	47819090	47819795	5730593F17Rik	0	chr17:47806578-47820892
chr17	56355449	56355594	Mpo	57556	chr17:56413150-56417783
chr17	79419543	79420749	Bahcc1	13038	chr17:79433787-79439686
chr17	80052124	80053059	Fasn	862	chr17:80053921-80071409
chr5	34655547	34655935	Rai14	301318	chr5:34957253-34958750
chr5	34687673	34687965	Rai14	269288	chr5:34957253-34958750
chr5	66491448	66492361	Cd180	-981	chr5:66488909-66490467
chr5	67585536	67585875	Pik3r1	0	chr5:67509819-67590624
chr5	77803248	77803367	Lhfpl2	0	chr5:77799632-77874486
chr5	106908226	1.07E+08	Efna5	0	chr5:106905355-106931655

chr5	108085570	1.08E+08	Fert2	0	chr5:108082660-108086010
chr5	148208221	1.48E+08	Adrb2	0	chr5:148204240-148255435
chr5	149464925	1.49E+08	Csf1r	25190	chr5:149490805-149493891
chr5	169760638	1.7E+08	Kcnmb1	-564	chr5:169756410-169760074
chr5	178803895	1.79E+08	Adamts2	-49139	chr5:178691630-178754756
chr7	1408433	1408951	Micall2	89119	chr7:1498070-1502021
chr7	26460385	26460633	Snx10	-20840	chr7:26437305-26439545
chr7	27201206	27201499	Hoxa9	0	chr7:27199735-27220891
chr7	27202553	27202993	Hoxa9	0	chr7:27199735-27220891
chr7	27203073	27203466	Hoxa9	0	chr7:27199735-27220891
chr7	50345715	50346835	Ikzf1	-235589	chr7:50107911-50110126
chr7	73624879	73625346	Lat2	-14413	chr7:73587401-73610466
chr7	76828348	76828576	Fgl2	16610	chr7:76845186-76846431
chr7	106704715	1.07E+08	Prkar2b	0	chr7:106618925-106721216
chr7	139529064	1.4E+08	Tbxas1	-51336	chr7:139471181-139477728
chr14	75361059	75361571	Dlst	-1825	chr14:75347589-75359234
chr14	77791405	77791671	Gstz1	-8940	chr14:77772242-77782465
chr14	89707672	89708616	Foxn3	-13657	chr14:89686978-89694015
chr14	93117257	93117768	Rin3	-63158	chr14:93052381-93054099
chr14	100531888	1.01E+08	Evl	9386	chr14:100542022-100544258
chr9	14203697	14204081	Nfib	0	chr9:14132520-14325750
chr9	32457089	32457364	Ddx58	-76	chr9:32383078-32457013
chr9	95726604	95726801	Fgd3	-25224	chr9:95690523-95701380
chr9	114424532	1.14E+08	Gng10	0	chr9:114409522-114425974
chr9	116379329	1.16E+08	Rgs3	0	chr9:116348945-116398471
chr9	117692618	1.18E+08	Tnfsf8	0	chr9:117690537-117702171
chr9	131369913	1.31E+08	Spna2	0	chr9:131369112-131371185
chr13	36049213	36050887	Mab21l1	0	chr13:36044089-36051808
chr13	40943529	40943850	Foxo1	0	chr13:40940889-40958355
chr13	46751279	46752608	Lcp1	-5166	chr13:46743863-46746113
chr13	46753288	46754147	Lcp1	-7175	chr13:46743863-46746113
chr13	50204794	50205145	Arl11	-4181	chr13:50189809-50200613
chr3	9911613	9911971	Cidec	0	chr3:9902655-9941408
chr3	24869934	24870287	Rarb	-186244	chr3:24680201-24683690
chr3	112692217	1.13E+08	Cd200r1	42134	chr3:112735476-112739663
chr3	112964544	1.13E+08	Boc	339	chr3:112965450-112966583
chr3	150322081	1.5E+08	2810407C02Rik	0	chr3:150320270-150323270
chr3	157325748	1.57E+08	Veph1	-33016	chr3:157290209-157292732
chr16	16123803	16124667	Abcc1	0	chr16:16107968-16124361
chr16	50715116	50715299	9130017C17Rik	13526	chr16:50728825-50732873
chr16	55540360	55540833	Mmp2	-17985	chr16:55504207-55522375
chr16	56137894	56138256	Gnao1	-266909	chr16:55869127-55870985
chr16	69457347	69457490	Cyb5b	-5180	chr16:69417593-69452167

chr16	87903345	87904117	BC048644	47037	chr16:87951154-87953835
chr16	88770501	88770848	Rnf166	0	chr16:88766514-88770583
chr21	36398548	36398828	Runx1	20276	chr21:36419104-36424477
chr21	36420456	36421805	Runx1	0	chr21:36419104-36424477
chr21	39868683	39869555	Erg	0	chr21:39830852-39877989
chr18	56809110	56809370	Sec11c	-86050	chr18:56708840-56723060
chr19	15574480	15574831	A430107D22Rik	-10634	chr19:15562049-15563846
chr19	40750567	40750963	Akt2	0	chr19:40734674-40793531
chr19	49841249	49841895	Cd37	493	chr19:49842388-49845001
chr19	51645546	51714605	Cd33	-3881	chr19:51631167-51641665
chr19	54876429	54876765	Lair1	0	chr19:54870067-54886694
chr11	6267881	6268720	Cnga4	-19648	chr11:6239734-6248233
chr11	10316893	10317243	Adm	0	chr11:10312838-10351582
chr11	10324978	10325218	Adm	0	chr11:10312838-10351582
chr11	47415025	47415403	Sfpi1	0	chr11:47412747-47423042
chr11	63528336	63528808	2700081O15Rik	5253	chr11:63534061-63538555
chr11	66673568	66673798	Pcx	0	chr11:66641441-66703669
chr11	67036383	67036711	Adrbk1	0	chr11:67029927-67057696
chr11	67165584	67166226	Ppp1ca	-23002	chr11:67138749-67142582
chr11	73692113	73692198	Ucp2	-2531	chr11:73687719-73689582
chr11	117858523	1.18E+08	Il10ra	0	chr11:117857980-117859426
chr11	122849412	1.23E+08	Bsx	-44045	chr11:122803497-122805367
chr15	37173455	37174113	Meis2	210804	chr15:37384917-37395168
chr15	61136559	61136807	Rora	-42316	chr15:61091049-61094243
chr15	96902600	96903113	Nr2f2	-1250	chr15:96867016-96901350
chr15	96905568	96906197	Nr2f2	-4218	chr15:96867016-96901350
chr15	96907528	96907914	Nr2f2	-6178	chr15:96867016-96901350
chr20	18269464	18269677	6330439K17Rik	31641	chr20:18301318-18303522
chr20	25033629	25034262	Acss1	0	chr20:25013038-25040024
chr20	25251494	25252151	Pygb	831	chr20:25252982-25254488
chr20	39314354	39315324	Mafb	0	chr20:39309302-39319699
chr20	47803610	47804113	Stau1	0	chr20:47801190-47805535
chr20	48924232	48924768	Cebpb	0	chr20:48860807-48930471
chr20	56135219	56135721	Pck1	0	chr20:56118475-56165809
chr20	56135935	56137320	Pck1	0	chr20:56118475-56165809
chr20	56138648	56138937	Pck1	0	chr20:56118475-56165809
chr20	56139464	56139982	Pck1	0	chr20:56118475-56165809
chr4	38858508	38859749	Tlr6	0	chr4:38857521-38860641
chr4	84034389	84034401	Plac8	-99147	chr4:83930601-83935242
chr4	108928564	1.09E+08	Hadh	0	chr4:108897305-108937926
chr4	113861384	1.14E+08	Larp7	-15696	chr4:113844242-113845688
chr4	184241311	1.84E+08	Cldn22	75490	chr4:184317353-184318852
chr4	184774713	1.85E+08	Stox2	-21381	chr4:184749297-184753332

chr4	185303774	1.85E+08	Irf2	0	chr4:185302924-185340443

Table 11**. Overlapping methylation change and adipose enhancer regions.** This table displays the 171 cross-species conserved and directionally consistent regions with differential methylation along with the nearest enhancer found in adipose tissue.

Table 12

Table 12. Human Subject Information

Obesity	Type 2 Diabetes	Age at surgery	BMI	Waist circumference (cm)
Ob	T2D	42	37.00	118
Ob	T2D	62	36.00	143
Ob	ND	43	40.30	137
Ob	T2D	56	37.10	124
Ob	ND	36	41.60	125
Ob	T2D	52	42.90	140
Ob	ND	49	40.00	140
Ob	T2D	41	40.40	125
Ob	ND	50	35.60	122
Ob	T2D	34	40.30	139
Ob	T2D	37	42.70	139
Ob	T2D	36	45.84	143
Ob	ND	41	37.22	122
Ob	ND	59	37.24	120
L	ND	39	24.50	87
L	ND	41	27.80	89
L	ND	59	25.10	100
L	ND	36	26.30	99
L	ND	39	28.40	102
L	ND	49	22.60	80
L	ND	41	20.00	

pre-op HDL (mol/L)	pre-op LDL (mol/L)	pre-op Cholesterol (mol/L)	Preop HbA1C mmol/mol
1	1.3	2.9	48
0.6	2	3.6	43
2	3.4	5.8	39
0.9		4	71
1.1	3.7	5.5	34
0.9	3.7	5.2	52
1.4	3.8	6.1	40
3.9	3.7	5	69
0.4	2.3	3.9	41
1	4.9	6.9	48
0.7	1.7	3.6	49
0.6	3.6	4.9	
1.1	4.4	6.2	4.4

1	3.1	4.9	
1.1	3.9	6	38
0.9	3.1	4.6	37
1.3	4.7	6.5	36
1.4	3.6	5.3	41
1	3	4.6	37
1.5	3.7	5.7	38

pre-op HbA1c (%)	pre-op glucose (mmol/L)	insulin pre-op (pg/ml)	insulin pre-op (pmol/L)
5.6	6.7		
5.2	6.7	1139.035	196.1148416
4.8	5.5	721.27	124.1856061
7.8	17.3	893.445	153.830062
4.3	5.4	1068.2	183.9187328
6	7.5	973.01	167.52927
4.9	5.9	579.32	99.74517906
7.6	8.2	850.8	146.4876033
5	5.3	893.095	153.7698003
5.6	7.9	1066.505	183.6268939
5.7	8.2	1347.56	232.0179063
5.6	6.6	1740.36	299.6487603
5.8			
4.2	5.2	1160.125	199.7460399
4.7	5.9	87.675	40
4.6	5.7	561.845	189
4.5	4.8	261.13	62.3
5	5.3	375.42	64.6
4.6	5.5	363.32	62.6
4.7	5.9	240.46	41.4
			47.2

HOMA-IR pre-op	preop TG (mol/L)	insulin 2 days after RYGB (pmol/L)	HOMA-IR 2 days after RYGB
	1.2		
45.32431895	3.3	118.612259	1033.849972
26.49292929	0.9	140.6043388	776.0460015
53.32775482	6.1	114.6969697	784.1707537
35.14891338	1.6		
44.67447199	1.2	182.0532025	1355.521783
21.72228344	2	100.1153581	443.8233033
49.48025712	1.4	131.0347796	853.109814

34.17106673	**2.6**		
45.7026936	**2.2**	**227.0377066**	**1852.899061**
58.77786961	**2.5**	**178.3384986**	**1839.010003**
74.57924702	**1.5**	**177.1935262**	**2359.814242**
37.28592746	**1.9**	**241.5693871**	**2144.556819**
8.4	**2.2**		
38.6	**1.3**		
12.5	**1**		
14.4	**0.92**		
12.8	**1.3**		
8.6	**1**		

follow up time (month)	Weight follow-up (kg)	BMI follow-up	Waist follow up (cm)
8	**128.1**	**31.4**	**129**
7	**112.5**	**31.5**	
7	**87.9**	**28.4**	**100**
6	**115.2**	**34.8**	**119**
5	**112**	**32**	**112**
7	**98.8**	**31.9**	**118**
9	**96.2**	**30**	**108**
5	**127.8**	**35.4**	**122**

Table 12**. Human Subject Information.** This table displays relevant information about the human subjects examined in this study.

Table 13

Table 13. **Pyrosequencing Primers**	
Runx1 Long 5'	**TTGAGTTTGTTAAATTTAGGGGTAAGT**
Runx1 Long 3'	**TTCAAACAACATTTTTAAATCATTTC**
Runx1 Nested 5'	**TTGAGTTTGTTAAATTTAGGGGTAAGT**
Runx1 Nested 3'	**/5Biosg/CAAACAAAAACTATTAACTCTAAAACCAC**
Runx1 Sequencing 1	**GTTATTAGAGTAAGTGTA**
Runx1 Sequencing 2	**GATAAAATTTAAAGAGTGTT**
Runx1 Sequencing 3	**GTTTAAGTATTTATTAAAATAT**
Pscdbp Long 5'	**TTTTTTAGGGAAAAGAATTTTTTTT**
Pscdbp Long 3'	**ATCTCAAACCACAACAATCACATAA**
Pscdbp Nested 5'	**GGTATTATTTTTAGTAGAGGTTAGGTAAA**
Pscdbp Nested 3'	**/5Biosg/CAAAACCAAAAAACCATATATTAAC**
Pscdbp Sequencing 1	**GGTATTATTTTTAGTAGAGGTTAGGTA**
Pscdbp Sequencing 2	**ATAGTGTTTGTTAGTATTTGAAT**
Stap1 Long 5'	**TTTTATAATAATTGAAGGAGGGAAAGT**
Stap1 Long 3'	**AAAAACAATATAACCCAAACAAAAAC**
Stap1 Nested 5'	**TTTTATAATAATTGAAGGAGGGAAAGT**
Stap1 Nested 3'	**/5Biosg/TACACACTTACCTAATAATCAAACC**
Stap1 Sequencing 1	**GAATAGTTGTTTTTTTTTTTAATAT**
Stap1 Sequencing 2	**AATTTTAAAGTAAAGGTTATG**
Pck1 Long 5'	**TGGTAAAGGTTTTGTTGTTTAAGTGT**
Pck1 Long 3'	**ATTCTCTAACCATCCCAAAATAAAC**
Pck1 Nested 5'	**TTTTTAGATATTTGGGTATTTAAGA**
Pck1 Nested 3'	**/5Biosg/ACTATAAACTTTATTCTAACAAAACAATAC**
Pck1 Sequencing 2	**GTTTGATGTATATTTTTTTG**
Pck1 Sequencing 3	**TTTAGAGTAGGGGTTAGTAT**
Dguok Long 5'	**AGTTGTATAATTTATTGTGGGTTGG**
Dguok Long 3'	**TCCTTTAAAACTTCCTCAAACATTT**
Dguok Nested 5'	**AGTTATGGAGGGTTAATTTGTGTTT**
Dguok Nested 3'	**/5Biosg/TCCTTTAAAACTTCCTCAAACATTT**
Dguok Sequencing 1	**GTTAATTTGTGTTTGGGATT**

Dguok Sequencing 2	**ATTTTTAGAGGATTGGGTGAAG**
Dguok Sequencing 3	**GGAGTTATTGATAAATTT**
Kcnj11 Long 5'	**AGAGTTTAGGTTATAGGTGGGAGGT**
Kcnj11 Long 3'	**AAAATATCCTACCAACCAAAAAAAA**
Kcnj11 Nested 5'	**TTTAGGTTATGTTTAAGGGTTTTGG**
Kcnj11 Nested 3'	**/5Biosg/AAAACTAAAAAACCCACACAAACAC**
Kcnj11 Sequencing 1	**ATTTTTTTTTTTAGTAATTTAGATAAG**
Kcnj11 Sequencing 2	**ATTTTATTTTTTAGTTTTTGG**
Kcnj11 Sequencing 3	**GAGGTGAGTTTAGGTAGATTT**
Zfhnx2 Long 5'	**TTTTGAGGGTTTAGTGGTAGTTTGT**
Zfhnx2 Long 3'	**CAAATTCAATAAAAACACAAATTAAAAAA**
Zfhnx2 Nested 5'	**AGGGTTTAGTGGTAGTTTGTTAGGTTAT**
Zfhnx2 Nested 3'	**/5Biosg/AAAAAAAATCAACAAAAACAATATTAAATT**
Zfhnx2 Sequencing 1	**AAATTGTTTATTTTTTGTAG**
Zfhnx2 Sequencing 2	**TTTTTTATTTTATTTTATTTTTT**
Slc38a4 Long 5'	**GGAGTTTTTATTAGGAGAGGGTGTAG**
Slc38a4 Long 3'	**ATAAACCAAAATAACCTCCAACTCA**
Slc38a4 Nested 5'	**TATTTTGATGGTGTTAGAGATGAATTT**
Slc38a4 Nested 3'	**/5Biosg/CCACCACCACCTAAAAAAAACT**
Slc38a4 Sequencing 1	**ATGGTGTTAGAGATGAATT**
Slc38a4 Sequencing 2	**GTATAGTTATATTTTTAAAT**
Slc38a4 Sequencing 3	**TTATTTGATGTTTAATATATTT**
Chpt1 Outer F	**GGGAGTTTAGAATAGTGATTGGTTG**
Chpt1 Outer R	**TCCCTTCCTTAAATAACCTTCCTAC**
Chpt1 Inner F	**GGGAAGATTTTTAAATAGAGAGGATGT**
Chpt1 Outer R	**/5BiosG/TACCCTAAAAACTAAACCCCAAAAC**
Chpt1 Seq2	**GATATTTAATATAATATTTATTT**
Chpt1 Seq3	**ATTGAGTTTTAGGTTAGTTTTA**
Chpt1 Seq4	**TTGGGGGAATTAATTTAAGT**
Liph Outer F	**ATAGTTTAGGAGGAAAGTTGGTGTT**

Liph Outer R	**AAAAACTTCCAAAACATAAAAAAAA**
Liph Inner F	**AAGGGATTTAAGGGATTTTTAAATTT**
Liph Inner R	**/5BiosG/CAACTCCCTACAACAAAACACTTT**
Liph Seq1	**GGTAGTTTGGTTTATTT**
Liph Seq2	**GATTATTTAGATGTTTGT**
Liph Seq3	**GTGGTATGATAGGTAGTAG**
Fasn Outer F	**GGGTTTTAAGAGGTTGTTGGTTAAT**
Fasn Outer R	**AACCCTAAACAAAACACAATTTCAC**
Fasn Inner F	**TAGTTTGTTTTGGGATAGGTTGTG**
Fasn Inner R	**/5BiosG/AATATACCCCCAAAAATAAAAAACC**
Fasn Seq1	**GAATATAAAGGTTAAGTGTTTA**
Fasn Seq2	**AGAGTTTGGGTAGTTAGATAG**
Fasn Seq3	**AGGGTTTTTATTTTTATTAAG**
Fasn Seq4	**GTTAATTAAATTTTTTAATTTG**
Axin2 Outer F	**GATTTTTAAGGAGGGGATTTTGTAG**
Axin2 Outer R	**CTATCCCACATCACCAATCTAAACT**
Axin2 Inner F	**TTTTTTTATTTTATGTGGTGTTTGT**
Axin2 Inner R	**/5BiosG/ACTAAAACTATCCCTACCTATTCCTC**
Axin2 Seq1	**TTTTATTTTATGTGGTGTT**
Axin2 Seq2	**TAATTGTTTTTGTTTTTTG**
Axin2 Seq3	**GAATTTTAGAGTGAGGATTTG**
Scd1 Outer F	**AGAAGGTTTGGGGTAATATAGAAGTTT**
Scd1 Outer R	**AAATCCCCTCTCCTTAAAACATAAC**
Scd1 Inner F	**TGGGAAATTTTTTGATAGTT**
Scd1 Inner R	**/5Biosg/AATTCTACTAAATCCTCAAAAAAACTAAAC**
Scd1 Seq1	**TTTGGGTTATATATGTGTTA**
Scd1 Seq2	**GTTTTTGTATTTGTGAGGG**
Scd1 Outer F	**AAGGGAGGTTTTTGTTATTTATTTA**
Scd1 Outer R	**ACCTTCCTTATAACCATCAATTCC**
Scd1 Inner F	**TGTTTTTTAGTAAGTGAGAAGAGATGGT**
Scd1 Inner R	**/5Biosg/AACCAAATTTAAACCCAACCTAAAC**
Scd1 Seq1	**GTGGTTTAGAAAGAAGAGTTTTGT**
Fermt2 Outer F	**TGTGGTAAGGTTTATTTTTTAGAGG**
Fermt2 Outer R	**AAACAAACTTTTATCTCCCCTTTTAC**
Fermt2 Inner F	**ATAATAGAGGATAGAAATAAAAGAATGAAA**
Fermt2 Inner R	**/5Biosg/CCTTCAAATTATATAAATTTCAAATAATAA**
Fermt2 Seq1	**ATTTTAGAAGTTATAATTTAAGTA**

Atp6v0a1 Outer F	**TTTTGGAATTAGTTTAAAAGGGTTG**
Atp6v0a1 Outer R	**AACAAAAAACAAAACAAAACCAAAT**
Atp6v0a1 Inner F	**ATTTTTAAGTAGGGATTTTTTTGTGAG**
Atp6v0a1 Inner R	**/5Biosg/TCATCCTAATACCTTCAAACTACTCTC**
Atp6v0a1 Seq1	**GATAAGATTTTAATGTATTTAAGTT**
Arhgap29 Outer F	**TATGGATTTTGGGATTTTTGATTAT**
Arhgap29 Outer R	**TACAACAACCTAACCAACAAAAAAA**
Arhgap29 Inner F	**TTTTAATTTTGAATTTAGAGGAAATTTAGT**
Arhgap29 Inner R	**/5Biosg/AAAATATTTAATAAATTTCTATTCCCCC**
Arhgap29 Seq1	**GTTATTTGTATTTTTGTTAAATT**
Arhgap29 Seq2	**TTGTAAAGTGTTTGTTGATAA**
Arhgap29 Seq3	**ATTAGATATTTTTGTTATAATTT**
Masp1 Outer F	**TATGTGTATTTATATTTGGGATTTTTTTAG**
Masp1 Outer R	**AATAAAACTCTTCTAACCCCTAAACTC**
Masp1 Inner F	**GAAGTTTGTGTTGTTTGTATTTTTG**
Masp1 Inner R	**/5Biosg/TCTATATTTACTTAAAACATACCCTC**
Masp1 Seq1	**GGTTTTGGTGTTTTTGGAGTGGGAGA**
Masp1 Seq2	**GTTATGTAGTAGGTGAAATGAGTT**
Elovl5 Outer F	**TTGTTGTATAGGTTTGTAGTTTAGGAGTAA**
Elovl5 Outer R	**TCAAAAACCCAATTAAATCAATATTC**
Elovl5 Inner F	**AAAATTGTTTAAGAGTATATTTTTTAAAAA**
Elovl5 Inner R	**/5Biosg/AACATCCAACATTAATTTCCTTACC**
Elovl5 Seq1	**ATTTTTATTGTTAAGTTATATATGATT**
Elovl5 Seq2	**TAGAGGTTTTTTAAAGAATGTG**
Elovl5 Seq3	**ATTTTGTAAATTAAATTAGATGG**
Tmem140 Outer F	**TATAGAATGATGTTTATAAAGTGGGATATA**
Tmem140 Outer R	**AATTAAAAACCCCATAACACTCTTCT**
Tmem140 Inner F	**GGGATATAAATATATATATTTATGTAATTT**
Tmem140 Inner R	**/5Biosg/AACCAATATTTCCTTCAAAAAACAA**
Tmem140 Seq1	**AGAAGTGTGTTGTTTAGAGTGGT**

Table 13. **Pyrosequencing Primers.** This table shows the primers that were used in our pyrosequencing assays

Table 14

Table S11. qPCR primers	
Tnks1bp1 F	CCCAGGACCCTCACTCCAT
Tnks1bp1 R	TCCCAAACTCCCAGTCTTGAA
Fbxw8 F	GCCAGGTTGCCTTTGGAGT
Fbxw8 R	TCCCGGATGTTGACACAGGTA
Sorbs1 F	CCCCGTCTGAGGTAATAGTTGT
Sorbs1 R	GAGCAGTCTCCAGGAGTATAGTC
Vps13c F	GAAGCTAAAGTAAAAGCCCACGA
Vps13c R	ACACATCAGAGGTGTTGACAATG
Tcf7l2 F	AACGAACACAGCGAATGTTTCC
Tcf7l2 R	CACCTTGTATGTAGCGAACGC
Pcx F	CTGAAGTTCCAAACAGTTCGAGG
Pcx R	CGCACGAAACACTCGGATG
Tnfsf8 F	GCACAAGTCGCAGCTACTTCT
Tnfsf8 R	GGAGTGGAGTCCTTTTTCTGG
Etaa1 F	GGTGGCACGGGAATGAGTC
Etaa1 R	GATTTGTACTGGCGTCTCCTTT
Pck1 F	CTGCATAACGGTCTGGACTTC
Pck1 R	CAGCAACTGCCCGTACTCC
Rgs3 F	GCTTCCTGTAGGACAAGACCT
Rgs3 R	GGCTTTGAGGGGGCTTAGG
Stau1 F	GGACCCTCACTCTCGGATG
Stau1 R	TTCTGGCAGGGGTTCACTCT
As3mt F	GGGAATGTACTGAAGACATCTGC
As3mt R	CCACAGCCATAATACCTCGAACT
Akt2 F	ACGTGGTGAATACATCAAGACC
Akt2 R	GCTACAGAGAAATTGTTCAGGGG
Tnfaip8l2 F	TCAGCTCAAAGAGTCTGGCAC
Tnfaip8l2 R	GGTAAAGCTCGTCTAGCACCTC

Table 14**. qPCR Primers.** This table lists the primers that were used in our quantitative PCR assays.

Table 15

Phenotype	Partitioning	Original Cohort	Post-QC removal	Paired Cohort
Total Subjects		524	459	346
Sex				
	Male	364	324	256
	Female	160	135	90
Age				
	Min	46	46	4
	1st Quartile	64.75	64	64
	Median	71	71	71
	3rd Quartile	77	77	77
	Max	90	88	86
BMI				
	Min	16.85	16.85	18.42
	1st Quartile	25.58	25.62	25.78
	Median	28.18	28.23	28.26
	3rd Quartile	31.36	31.42	31.6
	Max	47.26	47.26	45.97
Education				
	Grade 12, GED, or less	184	165	123
	Post-high school education, no bachelor's	161	141	100
	Bachelor's or graduate school degree	170	146	116
	Prefer not to answer	9	7	7
Smoking				
	Never smoked	129	114	87
	Former smoker	365	319	241
	Current smoker	30	26	18
Alcohol				
	No drinking	142	124	86
	Less than 7 drinks per week	238	215	167
	Greater than six drinks per week	142	118	92
Diabetes				
	No	413	370	275
	Yes	103	82	65
	Maybe	5	4	3

Table 15**. Univariate population statistics for Framingham Heart Study coronary heart disease cohort.**

Table 16

	FHS CAD cohort post-QC removals			Paired cohort (only paired samples)		
Phenotype	**Cases**	**Controls**	**p-value**	**Cases**	**Controls**	**p-value**
% Male	**74.90%**	**67.10%**	**0.4933**	**74%**	**74%**	**1**
Age	**71 (64,78)**	**71 (64,77)**	**0.9332**	**71 (64,78)**	**71 (64,77)**	**0.9949**
BMI	**28.53 (26.17,32.26)**	**28.02 (25.16,30.81)**	**0.1478**	**28.24 (26.05,32.35)**	**28.35 (25.33,31.24)**	**0.344**
Education					**5.3**	**0.0095**
Grade 12, GED, or less	**83**	**82**		**69**	**54**	
Less than Bachelor's	**61**	**80**	**0.0778**	**50**	**50**	**0.0449**
Bachelor's or graduate degree	**59**	**87**		**50**	**66**	
Smoking						
Never	**34**	**80**		**28**	**59**	
Former	**153**	**166**	**2.66E-06**	**130**	**111**	**5.28E-06**
Current	**20**	**6**		**15**	**3**	
Alcohol						
None	**58**	**66**		**47**	**39**	
Low	**90**	**125**	**0.632**	**80**	**87**	**0.549**
High	**59**	**59**		**46**	**46**	
High Blood Sugar or Diabetes						
Yes	**48**	**34**	**0.19**	**41**	**24**	**0.307**
No	**156**	**214**		**129**	**146**	

Table 16**. Bivariate statistics for the Framingham Heart Study coronary heart disease cohort**. Here we present population data for the final cohorts utilized in the CHD analysis, partitioning into cases and controls and analyzed to determine possible confounding effects of sensitivity variables versus the main case-control analysis. Binary variables were tested with Student's T-test. Ordinal variables were tested with linear trend tests.

Table 17

Probe	beta	pval	Gene	region	distance
cg24145146	-0.00755	7.56E-07	ZNF624	downstream	40208
cg03636183	-0.02817	2.97E-06	BTG2	inside	1263
cg03746015	-0.02237	4.71E-06	RGS9	upstream	14052
cg08417719	-0.00657	4.93E-06	CALCOCO1	promoter	100
cg15803765	0.00278	8.62E-06	CPD	promoter	478
cg03725309	-0.00964	9.71E-06	LINC00635	NA	0
cg21566642	-0.01942	1.08E-05	FBXL20	promoter	948
cg02741440	0.021064	1.45E-05	NMNAT3	inside	50151
cg08899667	0.015847	1.57E-05	UBE2G1	downstream	99148
cg18597709	0.011132	1.79E-05	SPRY4	promoter	457
cg10589385	0.041443	2.04E-05	ARRB2	inside	7445
cg19337279	0.017077	2.34E-05	MAN1C1	upstream	7282
cg00828616	-0.00466	2.63E-05	STAC3	inside	318
cg13078421	0.015352	2.80E-05	GUCY1B2	NA	0
cg05575921	-0.03299	2.86E-05	SYNGR3	inside	249
cg24878115	0.007848	2.99E-05	INS-IGF2	NA	0
cg17750139	-0.00234	3.11E-05	ZMYND15	inside	5639
cg23478225	-0.01344	3.90E-05	ARFGEF1	inside	88518
cg12347346	-0.02176	3.94E-05	TREML5P	NA	0
cg14172849	0.011574	4.19E-05	SPON1	inside	360
cg06756624	-0.00604	4.20E-05	MC1R	promoter	2019
cg05951221	-0.02419	4.28E-05	PCDH9	inside	128361

cg20309640	0.012297	4.39E-05	PITX2	downstream	11575
cg27388345	-0.00367	4.61E-05	AC007246.3	NA	0
cg12798040	0.013638	4.63E-05	GNG7	inside	155902
cg26575389	-0.01217	5.07E-05	ZNF175	promoter	61
cg11798876	-0.00634	5.32E-05	TANC2	inside	32270
cg26314512	-0.01269	5.49E-05	PDE4B	inside	225
cg21230425	-0.01775	5.52E-05	CHEK2P2	NA	0
cg04135110	0.01433	5.58E-05	KNDC1	inside	2013
cg04516896	-0.00674	5.65E-05	UBE2E2	inside	638
cg01940273	-0.02176	5.95E-05	PRKCZ	inside	1192
cg16717549	0.027613	6.22E-05	TCP11	inside	1126
cg04148762	0.002051	6.27E-05	YEATS2	promoter	173
cg16526137	-0.00331	6.80E-05	TBR1	inside	1062
cg14601444	0.015836	7.33E-05	POT1	NA	0
cg13126206	-0.00882	7.76E-05	LINC00937	NA	0
cg22382836	0.018655	7.93E-05	LOC100294145	NA	0
cg10673265	-0.00548	8.10E-05	FAM184B	inside	76
cg25512107	-0.00491	8.26E-05	CDKN1C	promoter	342
cg10752406	-0.00692	8.41E-05	TCERG1L	NA	0
cg25745600	-0.01096	8.59E-05	BPTF	promoter	93
cg09309711	0.006565	9.10E-05	GAK	inside	78939
cg23661483	0.011592	9.27E-05	RDH8	inside	8788
cg26507427	0.016834	9.55E-05	CHRND	inside	395
cg02170577	-0.00756	9.56E-	PRRC2A	inside	2420

		05			
cg27624229	-0.00808	9.61E-05	FAM46D	inside	29
cg25721516	-0.00789	9.78E-05	PEBP1	promoter	125
cg10209902	-0.00237	9.91E-05	PXDN	inside	66471

Table 17. **Top results of Framingham Heart Study coronary heart disease epigenome-wide association study.** Here we present the results of the primary model utilized to examine CHD. Probe represents the probe on the Illumina Infinium 450k BeadChip array and the corresponding CG. Beta represents the determined correlation coefficient between methylation and CHD status. p-value is the corresponding p-value for that coefficient. Gene is the nearest gene to that probe, and the region and distance columns represent the relationship of that probe to the gene.

Bibliography

Aapola, U., Kawasaki, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K.*, et al.* (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics *65*, 293-298.

Aapola, U., Liiv, I., and Peterson, P. (2002). Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. Nucleic acids research *30*, 3602-3608.

Almgren, P., Lehtovirta, M., Isomaa, B., Sarelin, L., Taskinen, M.R., Lyssenko, V., Tuomi, T., Groop, L., and Botnia Study, G. (2011). Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia *54*, 2811-2819.

Almind, K., and Kahn, C.R. (2004). Genetic determinants of energy expenditure and insulin resistance in diet-induced obesity in mice. Diabetes *53*, 3274-3285.

Ando, T., Nishimura, M., and Oka, Y. (2000). Decitabine (5-Aza-2'-deoxycytidine) decreased DNA methylation and expression of MDR-1 gene in K562/ADM cells. Leukemia *14*, 1915-1920.

Arner, P. (2003). The adipocyte in insulin resistance: key molecules and the impact of the thiazolidinediones. Trends Endocrinol Metab *14*, 137-145.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363-1369.

Aryee, M.J., Wu, Z., Ladd-Acosta, C., Herb, B., Feinberg, A.P., Yegnasubramanian, S., and Irizarry, R.A. (2011). Accurate genome-scale percentage DNA methylation estimates from microarray data. Biostatistics *12*, 197-210.

Bahl, J.J., Matsuda, M., DeFronzo, R.A., and Bressler, R. (1997). In vitro and in vivo suppression of gluconeogenesis by inhibition of pyruvate carboxylase. Biochem Pharmacol *53*, 67-74.

Bando, Y., Ushiogi, Y., Okafuji, K., Toya, D., Tanaka, N., and Fujisawa, M. (2001). The relationship of fasting plasma glucose values and other variables to 2-h postload plasma glucose in Japanese subjects. Diabetes Care *24*, 1156-1160.

Barres, R., Kirchner, H., Rasmussen, M., Yan, J., Kantor, F.R., Krook, A., Naslund, E., and Zierath, J.R. (2013). Weight loss after gastric bypass surgery in human obesity remodels promoter methylation. Cell Rep *3*, 1020-1027.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M.*, et al.* (2013). NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res *41*, D991-995.

Beale, E.G., Hammer, R.E., Antoine, B., and Forest, C. (2004). Disregulated glyceroneogenesis: PCK1 as a candidate diabetes and obesity gene. Trends Endocrinol Metab *15*, 129-135.

Bell, C.G., Finer, S., Lindgren, C.M., Wilson, G.A., Rakyan, V.K., Teschendorff, A.E., Akan, P., Stupka, E., Down, T.A., Prokopenko, I.*, et al.* (2010). Integrated Genetic and Epigenetic Analysis Identifies Haplotype-Specific Methylation in the FTO Type 2 Diabetes and Obesity Susceptibility Locus. PLoS One *5*.

Berry, M.N., and Friend, D.S. (1969). High-yield preparation of isolated rat liver parenchymal cells: a biochemical and fine structural study. J Cell Biol *43*, 506-520.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Human molecular genetics *9*, 2395-2402.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L.*, et al.* (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288-295.

Bjornsson, H.T., Fallin, M.D., and Feinberg, A.P. (2004). An integrated epigenetic and genetic approach to common human disease. Trends in genetics : TIG *20*, 350-358.

Boj, S.F., van Es, J.H., Huch, M., Li, V.S., Jose, A., Hatzis, P., Mokry, M., Haegebarth, A., van den Born, M., Chambon, P.*, et al.* (2012). Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. Cell *151*, 1595-1607.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nature reviews Genetics *10*, 295-304.

Cell_Signaling_Technology (2014). Histone Modification Table. In Cell Signaling Technology.

Cen, B., Selvaraj, A., Burgess, R.C., Hitzler, J.K., Ma, Z., Morris, S.W., and Prywes, R. (2003). Megakaryoblastic leukemia 1, a potent transcriptional coactivator for serum response factor (SRF), is required for serum induction of SRF target genes. Mol Cell Biol *23*, 6597-6608.

Chakrabarty, S., Nagata, M., Yasuda, H., Wen, L., Nakayama, M., Chowdhury, S.A., Yamada, K., Jin, Z., Kotani, R., Moriyama, H.*, et al.* (2003). Critical roles of CD30/CD30L interactions in murine autoimmune diabetes. Clin Exp Immunol *133*, 318-325.

Challen, G.A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z.*, et al.* (2014). Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. Cell stem cell *15*, 350-364.

Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S.J., and Park, J.H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet *45*, 400-405, 405e401-403.

Chedin, F., Lieber, M.R., and Hsieh, C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proceedings of the National Academy of Sciences of the United States of America *99*, 16916-16921.

Chen, L., Magliano, D.J., and Zimmet, P.Z. (2012). The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. Nat Rev Endocrinol *8*, 228-236.

Cho, H., Mu, J., Kim, J.K., Thorvaldsen, J.L., Chu, Q., Crenshaw, E.B., 3rd, Kaestner, K.H., Bartolomei, M.S., Shulman, G.I., and Birnbaum, M.J. (2001). Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta). Science *292*, 1728-1731.

Cohen, R.V., Pinheiro, J.C., Schiavon, C.A., Salles, J.E., Wajchenberg, B.L., and Cummings, D.E. (2012). Effects of gastric bypass surgery in patients with type 2 diabetes and only mild obesity. Diabetes Care *35*, 1420-1428.

Collado-Torres, L., and Jaffe, A.E. (2014). enrichedRanges: Identify enrichment between two sets of genomic ranges (GitHub).

Consortium, E.P., Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Consortium, G.L.G., Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L.*, et al.* (2013). Discovery and refinement of loci associated with lipid levels. Nat Genet *45*, 1274-1283.

Cooper, R., Cutler, J., Desvigne-Nickens, P., Fortmann, S.P., Friedman, L., Havlik, R., Hogelin, G., Marler, J., McGovern, P., Morosco, G.*, et al.* (2000). Trends and disparities in coronary heart disease, stroke, and other cardiovascular diseases in the United States: findings of the national conference on cardiovascular disease prevention. Circulation *102*, 3137-3147.

Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., Wu, Y., He, X., Powe, N.R., and Feinberg, A.P. (2003). Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. Science *299*, 1753-1755.

Dawber, T.R., Meadors, G.F., and Moore, F.E., Jr. (1951). Epidemiological approaches to heart disease: the Framingham Study. American journal of public health and the nation's health *41*, 279-281.

Dayeh, T., Volkov, P., Salo, S., Hall, E., Nilsson, E., Olsson, A.H., Kirkpatrick, C.L., Wollheim, C.B., Eliasson, L., Ronn, T.*, et al.* (2014). Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS genetics *10*, e1004160.

De Marcos Lousa, C., van Roermund, C.W., Postis, V.L., Dietrich, D., Kerr, I.D., Wanders, R.J., Baldwin, S.A., Baker, A., and Theodoulou, F.L. (2013). Intrinsic acyl-CoA thioesterase activity of a peroxisomal ATP

binding cassette transporter is required for transport and metabolism of fatty acids. Proc Natl Acad Sci U S A *110*, 1279-1284.

Del Guerra, S., Lupi, R., Marselli, L., Masini, M., Bugliani, M., Sbrana, S., Torri, S., Pollera, M., Boggi, U., Mosca, F.*, et al.* (2005). Functional and molecular defects of pancreatic islets in human type 2 diabetes. Diabetes *54*, 727-735.

Drobna, Z., Del Razo, L.M., Garcia-Vargas, G.G., Sanchez-Pena, L.C., Barrera-Hernandez, A., Styblo, M., and Loomis, D. (2013). Environmental exposure to arsenic, AS3MT polymorphism and prevalence of diabetes in Mexico. J Expo Sci Environ Epidemiol *23*, 151-155.

Dupont, C., Armant, D.R., and Brenner, C.A. (2009). Epigenetics: definition, mechanisms and clinical perspective. Seminars in reproductive medicine *27*, 351-357.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Fare, T.L., Coffey, E.M., Dai, H., He, Y.D., Kessler, D.A., Kilian, K.A., Koch, J.E., LeProust, E., Marton, M.J., Meyer, M.R.*, et al.* (2003). Effects of atmospheric ozone on microarray data quality. Analytical chemistry *75*, 4672-4675.

Feinberg, A.P., Irizarry, R.A., Fradin, D., Aryee, M.J., Murakami, P., Aspelund, T., Eiriksdottir, G., Harris, T.B., Launer, L., Gudnason, V.*, et al.* (2010). Personalized epigenomic signatures that are stable over time and covary with body mass index. Sci Transl Med *2*, 49ra67.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89-92.

Flanagan, S.E., Clauin, S., Bellanne-Chantelot, C., de Lonlay, P., Harries, L.W., Gloyn, A.L., and Ellard, S. (2009). Update of mutations in the genes encoding the pancreatic beta-cell K(ATP) channel subunits Kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism. Hum Mutat *30*, 170-180.

Funai, K., Song, H., Yin, L., Lodhi, I.J., Wei, X., Yoshino, J., Coleman, T., and Semenkovich, C.F. (2013). Muscle lipogenesis balances insulin sensitivity and strength through calcium signaling. J Clin Invest *123*, 1229-1240.

Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D*., et al.* (2010). A map of open chromatin in human pancreatic islets. Nat Genet *42*, 255-259.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307-315.

Gaziano, T.A., Bitton, A., Anand, S., Abrahams-Gessel, S., and Murphy, A. (2010). Growing epidemic of coronary heart disease in low- and middle-income countries. Current problems in cardiology *35*, 72-115.

Glass, J.L., Thompson, R.F., Khulan, B., Figueroa, M.E., Olivier, E.N., Oakley, E.J., Van Zant, G., Bouhassira, E.E., Melnick, A., Golden, A*., et al.* (2007). CG dinucleotide clustering is a species-specific property of the genome. Nucleic acids research *35*, 6798-6807.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395-398.

Gonzalez, Y., Herrera, M.T., Soldevila, G., Garcia-Garcia, L., Fabian, G., Perez-Armendariz, E.M., Bobadilla, K., Guzman-Beltran, S., Sada, E., and Torres, M. (2012). High glucose concentrations induce TNF-alpha production through the down-regulation of CD33 in primary human monocytes. BMC Immunol *13*, 19.

Gregor, M.F., and Hotamisligil, G.S. (2011). Inflammatory mechanisms in obesity. Annu Rev Immunol *29*, 415-445.

Griffith, J.S., and Mahler, H.R. (1969). DNA ticketing theory of memory. Nature *223*, 580-582.

Gual, P., Le Marchand-Brustel, Y., and Tanti, J.F. (2005). Positive and negative regulation of insulin signaling through IRS-1 phosphorylation. Biochimie *87*, 99-109.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D.*, et al.* (2011). Increased methylation variation in epigenetic domains across cancer types. Nature genetics *43*, 768-775.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nature cell biology *10*, 1291-1300.

Hansen, S.L., Myers, C.A., Charboneau, A., Young, D.M., and Boudreau, N. (2003). HoxD3 accelerates wound healing in diabetic mice. Am J Pathol *163*, 2421-2431.

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F.*, et al.* (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res *34*, D590-598.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947.

Holliday, R. (1987). The inheritance of epigenetic defects. Science *238*, 163-170.

Holliday, R. (1994). Epigenetics: an overview. Developmental genetics *15*, 453-457.

Hotamisligil, G.S. (2010). Endoplasmic reticulum stress and the inflammatory basis of metabolic disease. Cell *140*, 900-917.

Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. The Journal of biological chemistry *175*, 315-332.

Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86.

Huynh, J.L., Garg, P., Thin, T.H., Yoo, S., Dutta, R., Trapp, B.D., Haroutunian, V., Zhu, J., Donovan, M.J., Sharp, A.J*., et al.* (2014). Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. Nature neuroscience *17*, 121-130.

Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddeloh, J.A., Wen, B., and Feinberg, A.P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). Genome research *18*, 780-790.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M*., et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature genetics *41*, 178-186.

Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, M.D., Feinberg, A.P., and Irizarry, R.A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol *41*, 200-209.

Javierre, B.M., Fernandez, A.F., Richter, J., Al-Shahrour, F., Martin-Subero, J.I., Rodriguez-Ubreva, J., Berdasco, M., Fraga, M.F., O'Hanlon, T.P., Rider, L.G.*, et al.* (2010). Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome research *20*, 170-179.

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K.*, et al.* (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature *467*, 338-342.

Jin, W., Goldfine, A.B., Boes, T., Henry, R.R., Ciaraldi, T.P., Kim, E.Y., Emecan, M., Fitzpatrick, C., Sen, A., Shah, A.*, et al.* (2011). Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. J Clin Invest *121*, 918-929.

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., and de Bakker, P.I. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938-2939.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A.*, et al.* (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119-124.

Kanezaki, Y., Obata, T., Matsushima, R., Minami, A., Yuasa, T., Kishi, K., Bando, Y., Uehara, H., Izumi, K., Mitani, T.*, et al.* (2004). K(ATP) channel knockout mice crossbred with transgenic mice expressing a dominant-negative form of human insulin receptor have glucose intolerance but not diabetes. Endocr J *51*, 133-144.

Kannel, W.B., D'Agostino, R.B., and Belanger, A.J. (1987). Fibrinogen, cigarette smoking, and risk of cardiovascular disease: insights from the Framingham Study. American heart journal *113*, 1006-1010.

Kannel, W.B., Feinleib, M., McNamara, P.M., Garrison, R.J., and Castelli, W.P. (1979). An investigation of coronary heart disease in families. The Framingham offspring study. American journal of epidemiology *110*, 281-290.

Kashyap, S.R., Bhatt, D.L., Wolski, K., Watanabe, R.M., Abdul-Ghani, M., Abood, B., Pothier, C.E., Brethauer, S., Nissen, S., Gupta, M.*, et al.* (2013). Metabolic effects of bariatric surgery in patients with moderate obesity and type 2 diabetes: analysis of a randomized control trial comparing surgery with intensive medical treatment. Diabetes Care *36*, 2175-2182.

Khot, U.N., Khot, M.B., Bajzer, C.T., Sapp, S.K., Ohman, E.M., Brener, S.J., Ellis, S.G., Lincoff, A.M., and Topol, E.J. (2003). Prevalence of conventional risk factors in patients with coronary heart disease. Jama *290*, 898-904.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.*, et al.* (2010a). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285-290.

Kim, M., Long, T.I., Arakawa, K., Wang, R., Yu, M.C., and Laird, P.W. (2010b). DNA methylation as a biomarker for cardiovascular disease risk. PloS one *5*, e9692.

Kirchner, H., Nylen, C., Laber, S., Barres, R., Yan, J., Krook, A., Zierath, J.R., and Naslund, E. (2014). Altered promoter methylation of PDK4, IL1 B, IL6, and TNF after Roux-en Y gastric bypass. Surg Obes Relat Dis.

Kiryluk, K., and Isom, R. (2007). Thiazolidinediones and fluid retention. Kidney Int *72*, 762-768.

Ladd-Acosta, C., Aryee, M.J., Ordway, J.M., and Feinberg, A.P. (2010). Comprehensive high-throughput arrays for relative methylation (CHARM). Curr Protoc Hum Genet *Chapter 20*, Unit 20 21 21-19.

Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841-1842.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., and Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nature reviews Genetics *11*, 733-739.

Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS genetics *3*, 1724-1735.

Li, B.Z., Huang, Z., Cui, Q.Y., Song, X.H., Du, L., Jeltsch, A., Chen, P., Li, G., Li, E., and Xu, G.L. (2011). Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell research *21*, 1172-1181.

Li, L.C., and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. Bioinformatics *18*, 1427-1431.

Li, W.C., Ralphs, K.L., and Tosh, D. (2010). Isolation and culture of adult mouse hepatocytes. Methods Mol Biol *633*, 185-196.

Liu, C.T., Monda, K.L., Taylor, K.C., Lange, L., Demerath, E.W., Palmas, W., Wojczynski, M.K., Ellis, J.C., Vitolins, M.Z., Liu, S.*, et al.* (2013a). Genome-wide association of body fat distribution in African ancestry populations suggests new loci. PLoS genetics *9*, e1003681.

Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M.*, et al.* (2013b). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nature biotechnology *31*, 142-147.

Lloyd-Jones, D.M., Larson, M.G., Beiser, A., and Levy, D. (1999). Lifetime risk of developing coronary heart disease. Lancet *353*, 89-92.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y.*, et al.* (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet *380*, 2095-2128.

Lunnon, K., Smith, R., Hannon, E., De Jager, P.L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R.*, et al.* (2014). Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nature neuroscience *17*, 1164-1170.

Lynch, A.G., Chin, S.F., Dunning, M.J., Caldas, C., Tavare, S., and Curtis, C. (2012). Calling sample mix-ups in cancer population studies. PloS one *7*, e41815.

Lynch, C.J., Fox, H., Hazen, S.A., Stanley, B.A., Dodgson, S., and Lanoue, K.F. (1995). Role of hepatic carbonic anhydrase in de novo lipogenesis. Biochem J *310 (Pt 1)*, 197-202.

Marenberg, M.E., Risch, N., Berkman, L.F., Floderus, B., and de Faire, U. (1994). Genetic susceptibility to death from coronary heart disease in a study of twins. The New England journal of medicine *330*, 1041-1046.

Martin, B.C., Warram, J.H., Krolewski, A.S., Bergman, R.N., Soeldner, J.S., and Kahn, C.R. (1992). Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study. Lancet *340*, 925-929.

Matthews, K.A., Kelsey, S.F., Meilahn, E.N., Kuller, L.H., and Wing, R.R. (1989). Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. American journal of epidemiology *129*, 1132-1144.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell research *23*, 1256-1269.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y.*, et al.* (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253-257.

McCarthy, M.I. (2010). Genomics, type 2 diabetes, and obesity. N Engl J Med *363*, 2339-2350.

Migheli, F., Stoccoro, A., Coppede, F., Wan Omar, W.A., Failli, A., Consolini, R., Seccia, M., Spisni, R., Miccoli, P., Mathers, J.C.*, et al.* (2013). Comparison study of MS-HRM and pyrosequencing techniques for quantification of APC and CDKN2A gene methylation. PLoS One *8*, e52501.

Millward, C.A., Desantis, D., Hsieh, C.W., Heaney, J.D., Pisano, S., Olswang, Y., Reshef, L., Beidelschies, M., Puchowicz, M., and Croniger, C.M. (2010). Phosphoenolpyruvate carboxykinase (Pck1) helps regulate the triglyceride/fatty acid cycle and development of insulin resistance in mice. J Lipid Res *51*, 1452-1463.

Mingrone, G., Panunzi, S., De Gaetano, A., Guidone, C., Iaconelli, A., Leccesi, L., Nanni, G., Pomp, A., Castagneto, M., Ghirlanda, G.*, et al.* (2012). Bariatric surgery versus conventional medical therapy for type 2 diabetes. N Engl J Med *366*, 1577-1585.

Montano, C.M., Irizarry, R.A., Kaufmann, W.E., Talbot, K., Gur, R.E., Feinberg, A.P., and Taub, M.A. (2013). Measuring cell-type specific differential methylation in human brain tissue. Genome Biol *14*, R94.

Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segre, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A.*, et al.* (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet *44*, 981-990.

Multhaup, M.L., Seldin, M.M., Jaffe, A.E., Lei, X., Kirchner, H., Mondal, P., Li, Y., Rodriguez, V., Drong, A., Hussain, M.*, et al.* (2015). Mouse-human experimental epigenetic analysis unmasks dietary targets and genetic liability for diabetic phenotypes. Cell metabolism *21*, 138-149.

Nguyen, A., Rauch, T.A., Pfeifer, G.P., and Hu, V.W. (2010). Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *24*, 3036-3051.

Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS genetics *6*, e1000888.

Nilsson, E., Jansson, P.A., Perfilyev, A., Volkov, P., Pedersen, M., Svensson, M.K., Poulsen, P., Ribel-Madsen, R., Pedersen, N.L., Almgren, P.*, et al.* (2014). Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. Diabetes *63*, 2962-2976.

Noble, D. (2012). A theory of biological relativity: no privileged level of causation. Interface focus *2*, 55-64.

Oberdoerffer, S. (2012). A conserved role for intragenic DNA methylation in alternative pre-mRNA splicing. Transcription *3*, 106-109.

Ohlsson, R., Kanduri, C., Whitehead, J., Pfeifer, S., Lobanenkov, V., and Feinberg, A.P. (2003). Epigenetic variability and the evolution of human cancer. Advances in cancer research *88*, 145-168.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nature genetics *19*, 219-220.

Olszewski, P.K., Rozman, J., Jacobsson, J.A., Rathkolb, B., Stromberg, S., Hans, W., Klockars, A., Alsio, J., Riserus, U., Becker, L.*, et al.* (2012). Neurobeachin, a regulator of synaptic protein targeting, is associated with body fat mass and feeding behavior in mice and body-mass index in humans. PLoS genetics *8*, e1002568.

Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, İ., Tena, J.J., Morán, I., Gómez-Marín, C., van de Bunt, M.*, et al.* (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet.

Pearson, T.A., Blair, S.N., Daniels, S.R., Eckel, R.H., Fair, J.M., Fortmann, S.P., Franklin, B.A., Goldstein, L.B., Greenland, P., Grundy, S.M.*, et al.* (2002). AHA Guidelines for Primary Prevention of Cardiovascular Disease and Stroke: 2002 Update: Consensus Panel Guide to Comprehensive Risk Reduction for Adult Patients Without Coronary or Other Atherosclerotic Vascular Diseases. American Heart Association Science Advisory and Coordinating Committee. Circulation *106*, 388-391.

Prokunina-Olsson, L., Kaplan, L.M., Schadt, E.E., and Collins, F.S. (2009). Alternative splicing of TCF7L2 gene in omental and subcutaneous adipose tissue and risk of type 2 diabetes. PLoS One *4*, e7231.

Rahman, S.M., Janssen, R.C., Choudhury, M., Baquero, K.C., Aikens, R.M., de la Houssaye, B.A., and Friedman, J.E. (2012). CCAAT/enhancer-binding protein beta (C/EBPbeta) expression regulates dietary-induced inflammation in macrophages and adipose tissue in mice. J Biol Chem *287*, 34349-34360.

Rakyan, V.K., Beyan, H., Down, T.A., Hawa, M.I., Maslau, S., Aden, D., Daunay, A., Busato, F., Mein, C.A., Manfras, B.*, et al.* (2011). Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. PLoS genetics *7*, e1002300.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences of the United States of America *97*, 5237-5242.

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. Current opinion in genetics & development *12*, 162-169.

Ribel-Madsen, R., Fraga, M.F., Jacobsen, S., Bork-Jensen, J., Lara, E., Calvanese, V., Fernandez, A.F., Friedrichsen, M., Vind, B.F., Hojlund, K.*, et al.* (2012). Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. PLoS One *7*, e51302.

Roberts, R., Hodson, L., Dennis, A.L., Neville, M.J., Humphreys, S.M., Harnden, K.E., Micklem, K.J., and Frayn, K.N. (2009). Markers of de novo lipogenesis in adipose tissue: associations with small adipocytes and insulin sensitivity in humans. Diabetologia *52*, 882-890.

Rome, S., Meugnier, E., Lecomte, V., Berbe, V., Besson, J., Cerutti, C., Pesenti, S., Granjon, A., Disse, E., Clement, K.*, et al.* (2009). Microarray analysis of genes with impaired insulin regulation in the skeletal muscle of type 2 diabetic patients indicates the involvement of basic helix-loop-helix domain-containing, class B, 2 protein (BHLHB2). Diabetologia *52*, 1899-1912.

Ronn, T., Volkov, P., Davegardh, C., Dayeh, T., Hall, E., Olsson, A.H., Nilsson, E., Tornberg, A., Dekker Nitert, M., Eriksson, K.F.*, et al.* (2013). A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS genetics *9*, e1003572.

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., and Esteller, M. (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics *6*, 692-702.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421-427.

Schunkert, H., Konig, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C.*, et al.* (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nature genetics *43*, 333-338.

Sharma, P., Garg, G., Kumar, A., Mohammad, F., Kumar, S.R., Tanwar, V.S., Sati, S., Sharma, A., Karthikeyan, G., Brahmachari, V.*, et al.* (2014). Genome wide DNA methylation profiling for epigenetic alteration in coronary artery disease patients. Gene *541*, 31-40.

Sharma, P., Kumar, J., Garg, G., Kumar, A., Patowary, A., Karthikeyan, G., Ramakrishnan, L., Brahmachari, V., and Sengupta, S. (2008). Detection of altered global DNA methylation in coronary artery disease patients. DNA and cell biology *27*, 357-365.

Sharma, S., Kelly, T.K., and Jones, P.A. (2010). Epigenetics in cancer. Carcinogenesis *31*, 27-36.

Shen, J.C., Rideout, W.M., 3rd, and Jones, P.A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic acids research *22*, 972-976.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74-79.

Shuto, T., Furuta, T., Oba, M., Xu, H., Li, J.D., Cheung, J., Gruenert, D.C., Uehara, A., Suico, M.A., Okiyoneda, T.*, et al.* (2006). Promoter hypomethylation of Toll-like receptor-2 gene is associated with increased proinflammatory response toward bacterial peptidoglycan in cystic fibrosis bronchial epithelial cells. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *20*, 782-784.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article3.

Smyth, G.K. (2005). Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R.A. Irizarry, and W. Huber, eds. (New York: Springer), pp. 397-420.

Stahl, A., Evans, J.G., Pattel, S., Hirsch, D., and Lodish, H.F. (2002). Insulin causes fatty acid transport protein translocation and enhanced fatty acid uptake in adipocytes. Dev Cell *2*, 477-488.

Su, J., Lucchesi, P.A., Gonzalez-Villalobos, R.A., Palen, D.I., Rezk, B.M., Suzuki, Y., Boulares, H.A., and Matrougui, K. (2008). Role of advanced glycation end products with oxidative stress in resistance artery dysfunction in type 2 diabetic mice. Arterioscler Thromb Vasc Biol *28*, 1432-1438.

Surwit, R.S., Feinglos, M.N., Rodin, J., Sutherland, A., Petro, A.E., Opara, E.C., Kuhn, C.M., and Rebuffe-Scrive, M. (1995). Differential effects of fat and sucrose on the development of obesity and diabetes in C57BL/6J and A/J mice. Metabolism *44*, 645-651.

Surwit, R.S., Kuhn, C.M., Cochrane, C., McCubbin, J.A., and Feinglos, M.N. (1988). Diet-induced type II diabetes in C57BL/6J mice. Diabetes *37*, 1163-1167.

Tang, X., Guilherme, A., Chakladar, A., Powelka, A.M., Konda, S., Virbasius, J.V., Nicoloro, S.M., Straubhaar, J., and Czech, M.P. (2006). An RNA interference-based screen identifies MAP4K4/NIK as a negative regulator of PPARgamma, adipogenesis, and insulin-responsive hexose transport. Proc Natl Acad Sci U S A *103*, 2087-2092.

Tate, P.H., and Bird, A.P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Current opinion in genetics & development *3*, 226-231.

Therneau, T.M. (1999). A package for survival analysis in S (Mayo Foundaton).

Tost, J., and Gut, I.G. (2007). DNA methylation analysis by pyrosequencing. Nature protocols *2*, 2265-2275.

Virtanen, K.A., Lonnroth, P., Parkkola, R., Peltoniemi, P., Asola, M., Viljanen, T., Tolvanen, T., Knuuti, J., Ronnemaa, T., Huupponen, R.*, et al.* (2002). Glucose uptake and perfusion in subcutaneous and visceral adipose tissue during insulin stimulation in nonobese and obese humans. J Clin Endocrinol Metab *87*, 3902-3910.

Volkmar, M., Dedeurwaerder, S., Cunha, D.A., Ndlovu, M.N., Defrance, M., Deplus, R., Calonne, E., Volkmar, U., Igoillo-Esteve, M., Naamane, N.*, et al.* (2012). DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. The EMBO journal *31*, 1405-1426.

Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters. Nature *150*, 563-565.

Wang, K., Li, M., and Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. Nature reviews Genetics *11*, 843-854.

Wang, T., Pan, Q., Lin, L., Szulwach, K.E., Song, C.X., He, C., Wu, H., Warren, S.T., Jin, P., Duan, R.*, et al.* (2012). Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. Human molecular genetics *21*, 5500-5510.

Wang, X., and Seed, B. (2003). A PCR primer bank for quantitative gene expression analysis. Nucleic Acids Res *31*, e154.

Wang, Y.-H., Yu, Alice L. (2012). Abstract 3317: The roles of TMCC3 in breast cancer cells. Cancer Res *72*.

Wareham, N.J., Franks, P.W., and Harding, A.H. (2002). Establishing the role of gene-environment interactions in the etiology of type 2 diabetes. Endocrinol Metab Clin North Am *31*, 553-566.

Wei, L., Liu, S., Su, Z., Cheng, R., Bai, X., and Li, X. (2014). LINE-1 hypomethylation is associated with the risk of coronary heart disease in Chinese population. Arquivos brasileiros de cardiologia *102*, 481-488.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L.*, et al.* (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research *42*, D1001-1006.

West, D.B., Waguespack, J., and McCollister, S. (1995). Dietary obesity in the mouse: interaction of strain with diet composition. Am J Physiol *268*, R658-665.

Wray, N.R., Yang, J., Goddard, M.E., and Visscher, P.M. (2010). The genetic interpretation of area under the ROC curve in genomic profiling. PLoS genetics *6*, e1000864.

Xu, H., Barnes, G.T., Yang, Q., Tan, G., Yang, D., Chou, C.J., Sole, J., Nichols, A., Ross, J.S., Tartaglia, L.A.*, et al.* (2003). Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. J Clin Invest *112*, 1821-1830.

Xu, J., Wang, P., Li, Y., Li, G., Kaczmarek, L.K., Wu, Y., Koni, P.A., Flavell, R.A., and Desir, G.V. (2004). The voltage-gated potassium channel Kv1.3 regulates peripheral insulin sensitivity. Proc Natl Acad Sci U S A *101*, 3112-3117.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W*., et al.* (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet *42*, 565-569.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer cell *26*, 577-590.

Zarse, K., Schmeisser, S., Groth, M., Priebe, S., Beuster, G., Kuhlow, D., Guthke, R., Platzer, M., Kahn, C.R., and Ristow, M. (2012). Impaired insulin/IGF1 signaling extends life span by promoting mitochondrial L-proline catabolism to induce a transient ROS signal. Cell metabolism *15*, 451-465.

Zhang, L., Wang, Y., Xiao, F., Wang, S., Xing, G., Li, Y., Yin, X., Lu, K., Wei, R., Fan, J*., et al.* (2014). CKIP-1 regulates macrophage proliferation by inhibiting TRAF6-mediated Akt activation. Cell Res *24*, 742-761.

Zhang, S., Zhang, Y., Wei, X., Zhen, J., Wang, Z., Li, M., Miao, W., Ding, H., Du, P., Zhang, W*., et al.* (2010). Expression and regulation of a novel identified TNFAIP8 family is associated with diabetic nephropathy. Biochim Biophys Acta *1802*, 1078-1086.

Zhang, Y., Rohde, C., Tierling, S., Stamerjohanns, H., Reinhardt, R., Walter, J., and Jeltsch, A. (2009). DNA methylation analysis by bisulfite conversion, cloning, and sequencing of individual clones. Methods in molecular biology *507*, 177-187.

Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E*., et al.* (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477-481.

Curriculum Vitae

Michael L. Multhaup
222 Julia Ave, Mill Valley, California
(415) 577-0112; mmulthaup@gmail.com

Summary of Qualifications: Extensive experience in life science research, specializing in molecular biology and epigenetics. Multiple Big Data analysis projects, examining genome-wide data with millions of data points and hundreds of samples. Papers in Cell Metabolism and in press at Genome Biology and presentations at four national conferences.

Education

Johns Hopkins University School of Medicine, Baltimore, MD 2008-2015

- Doctor of Philosophy (PhD) in Molecular Biology and Genetics, scheduled January 2015
- Laboratory of Andrew P. Feinberg, M.D., Division of Molecular Medicine, 2009-2015
- Graduate program in Biochemistry, Cellular and Molecular Biology

Stanford University, Stanford, CA 2004-2008

- Bachelor of Arts (BA) in Biological Sciences, Molecular Biology Specialization, Spring 2008
- Graduated with Honors Thesis and Departmental Honors, 2008
- Stanford Summer Research Program, Laboratory Internship, mast cell biology, 2006-2007

Graduate Research Projects

Mouse-human analysis unmasks dietary targets and genetic liability for diabetic phenotypes

- Initiated and led collaboration with the Center for Metabolism and Obesity Research (Baltimore, MD) and the Department of Physiology Karolinska Institutet (Stockholm, Sweden)
- Used mouse models of insulin resistance and obesity to examine human Type 2 Diabetes
- Mouse care and surgery, transducted cell lines, human tissue samples, genome-wide analyses

Coronary artery disease and circulating DNA methylation in older adults

- Led collaboration with the Framingham Heart Study, Framingham, MA
- Epidemiologic analysis of epigenetic changes in Coronary Heart Disease
- Custom mixed-effect and logistic regression models coded with R

Publications

- First author: Multhaup et al. **Mouse-human experimental epigenetic analysis unmasks dietary targets and genetic liability for diabetic phenotypes.** Cell Metabolism, 2015.

- First author: Multhaup, et al. **Coronary artery disease and circulating DNA methylation in older adults.** In preparation.

● Riccardo E Marioni… **Michael L Multhaup**, et al. **DNA methylation age of blood predicts all-cause mortality in later life.** Genome Biology, in press.

Presentations at National Institutes of Health (NIH) sponsored scientific conferences

Animal Models and Personalized Medicine, NIH campus, Bethesda, MD 2013

● Invited Talk: Animal and human models of epigenetic dysregulation in metabolic dysfunction

Centers of Excellence in Genomic Sciences 2010, 2012, 2013

● Invited Talk: Conserved epigenetic changes in metabolic dysfunction (Madison, WI, 2013)

Biology Skills

● Wide range of undergraduate and graduate classes in statistics, history and theory of biology

● Molecular biology techniques including: cloning, Western blots, ELISAs, DNA and RNA extraction, cDNA synthesis, quantitative PCR, column chromatography, Sanger sequencing, bisulfite DNA treatment, pyrosequencing, lentiviral creation

● Cellular biology skills include primary and immortal cell culture, plasmid transfection, transduction and transfection, adipocyte and hepatocyte purification from whole tissue

● Previous mouse work has included general mouse care, surgical procedures to extract mouse blood, adipose tissue, liver, heart, pancreas, muscle and hypothalamus, fasting glucose tests and glucose and insulin tolerance tests, laser-eye surgery

Data Analysis Skills

● Genome-wide data: microarray preparation and next-generation sequencing (454, MiSeq).

● Utilized principal components (multi-dimensional scaling) and surrogate variable analysis to examine and remove technical batches in large datasets

● Found associations between genome-wide data and disease phenotypes using multiple types of regression models – linear, mixed-effect, and logistic regression

● All data analysis done using the programming language R. Extensive experience in both adapting R scripts written by others as well as writing entirely new code for novel uses

Leadership Experience

Johns Hopkins Graduate Student Association Executive Board, Baltimore, Md 2013-2014

● Nominated by fellow students, started as Secretary, later promoted to Vice-President

● Awarded and distributed funds for graduate student interest groups and individual travel awards

● Invited speakers and arranged seminars and lectures at Johns Hopkins

● Collaborated with faculty to establish recurring graduate student events

Teaching Assistant for Genetics Class, Baltimore, Md 2011-2014

● Taught epigenetics to first-year graduate students

● Developed in-depth exam questions with faculty and graded final examinations

Teacher Liaison with the Incentive Mentoring Program, Baltimore, Md 2008-2009

● Mentored at-risk high school students on general coursework and college applications

- Consulted with teachers for all mentored students to coordinate coursework and assignments

Professional Experience

Contract Editor with American Journal Experts 2012-2014

American Journal Experts edits technical papers written by non-native English-language scientists to a standard suitable for publication in American and English scientific journals.

- As a contract editor, I made thousands of edits across 42 scientific papers for publication in a wide variety of journals
- Editing included grammar, phrasing, fluency, and the correct usage of technical language

Other Skills:

- Can explain and present scientific concepts – more than a hundred data presentations
- Experienced scuba diver - can keep cool under pressure
- Badminton Singles Champion, San Francisco Bay League, 2004 – competitive spirit